

AL-TR-1991-0031

AD-A243 802



2

NEURAL NETWORKS AND THEIR APPLICATION TO
AIR FORCE PERSONNEL MODELING

Vince L. Wiggins

RRC, Incorporated
3833 Texas Avenue, Suite 256
Bryan, TX 77802

Larry T. Looper
Sheree K. Engquist, Capt, USAF

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

November 1991

Final Technical Report for Period September 1989 -- May 1991

Approved for public release; distribution is unlimited.

91-18549



AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

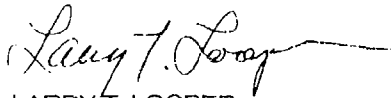
ARMSTRONG
LABORATORY

NOTICES

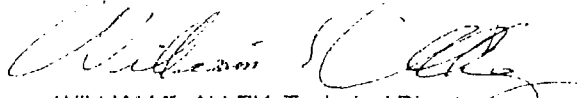
When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

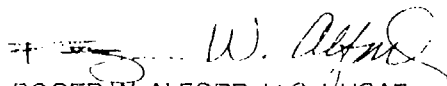
This report has been reviewed and is approved for publication.



LARRY T. LOOPER
Project Scientist



WILLIAM E. ALLEY, Technical Director
Manpower and Personnel Research Division



ROGER W. ALFORD, Lt Col, USAF
Chief, Manpower and Personnel Research Division

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE November 1991	3. REPORT TYPE AND DATES COVERED Final Report - September 1989 - May 1991	
4. TITLE AND SUBTITLE Neural Networks and Their Application to Air Force Personnel Modeling			5. FUNDING NUMBERS C - F41689-88-D-0251 PE - 62205F PR - 7719 TA - 20 WU - 20	
6. AUTHOR(S) Vince L. Wiggins Larry T. Looper Sheree K. Engquist				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) RRC, Incorporated 3833 Texas Avenue, Suite 256 Bryan, TX 77802			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Armstrong Laboratory Human Resources Directorate Manpower and Personnel Research Division Brooks Air Force Base, TX 78235-5000			10. SPONSORING/MONITORING AGENCY REPORT NUMBER AL-TR-1991-0031	
11. SUPPLEMENTARY NOTES Armstrong Laboratory Technical Monitor: Larry T. Looper, (512) 536-3648				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Neural network technology has recently demonstrated capabilities in areas important to personnel research such as statistical analysis, decision modeling, control, and forecasting. The present investigation indicates that three different neural network architectures are particularly suited to modeling many aspects of the Air Force personnel system: back propagation, learning vector quantization, and probabilistic neural networks. The primary advantage of neural networks is their ability to derive nonlinear and interacting relationships among model variables. Two areas investigated in order to evaluate this capability were airmen reenlistment decisions and airman inventory modeling.				
14. SUBJECT TERMS decision modeling neural networks personnel system modeling			15. NUMBER OF PAGES 64	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

CONTENTS

	Page
SUMMARY	1
INTRODUCTION	1
INTRODUCTION TO NEURAL NETWORKS	2
Artificial Neurons	3
Neural Network Architectures	3
Back Propagation	3
Training With Back Propagation	6
Capabilities of Back Propagation	8
Back Propagation Problems and Solutions	9
Learning Vector Quantization	13
Competitive Learning in LVQ	13
Related Architectures and Improvements	15
Probabilistic Neural Network	16
Overview	16
Training and Classification With a PNN	16
Other Architectures	20
AIR FORCE PERSONNEL MODELING	21
Types of Personnel Models	23
Accession/Enlistment Models	23
Aggregate Accessions	23
Applying Neural Networks to Aggregate Accessions	24
Individual Enlistment	24
Applying Neural Networks to Individual Enlistment	25
Reenlistment/Separation	25
Some Specific Models	25
Reenlistment Assessment	29
Neural Network Systems and Reenlistment Models	30
Prior Service	31
Inventory Planning Models (IPMs)	31
Cohort-Based Inventory Models	31
Other Inventory Models	32
Neural Networks and IPMs	33

CONTENTS (Continued)

	Page
Other Personnel-Related Models	34
Armed Forces Health Professions Scholarship Program (AFHPSP)	34
Neural Networks and the AFHPSP	34
Recruiter Assignments.	34
IMPLEMENTING NEURAL NETWORK PERSONNEL MODELS	35
Reenlistment Model	35
Model Structure	36
Modeling Techniques	36
Data Requirements	37
Validation and Testing	37
Evaluation and Interpretation of Models	38
Inventory Model	40
Initial Network Model	41
Validation	43
CONCLUSIONS	43
REFERENCES	44

List of Figures

Fig No.		Page
1	An artificial neuron with some reenlistment determinants as direct inputs	4
2	A simple back propagation network to predict reenlistment/ separation decisions of enlisted airmen	5
3	Schematic and computations for LVQ	14
4	A hypothetical distribution of airmen at a reenlistment/separation decision point and the decision regions formed by applying the LVQ architecture of Figure 3 to this distribution	15
5	Application of the Bayesian minimum loss decision rule using hypothetical distributions of airmen at a reenlistment decision point	18
6	Effect of changing the smoothing parameter σ on the form of an estimated PDF .	18
7	A conceptual view of the airmen and information flows in the enlisted personnel system	22

List of Tables

Table No.		
1	INDEPENDENT VARIABLES USED IN FIRST TERM REENLISTMENT/RETENTION MODELS	27
2	SIMULTANEOUS ACCESSION/RETENTION EQUATION SYSTEM	41

A-1

PREFACE

This is the first task in a two-stage effort to assess the potential for applying neural network techniques to the Air Force personnel field. The current work provides a conceptual overview of the technology and recommendations for specific application areas. The second task will directly assess the empirical capabilities of neural networks as compared to those of more traditional methods. These efforts are a component of the Armstrong Laboratory Force Management Program. The resulting models will serve as analysis and decision tools in the Air Force and OASD force management and policy analysis systems.

The authors wish to particularly thank Ms. Kathryn Turner for substantial revisions to Section II of this document. Ms. Amy Wortman provided many suggestions to improve the readability of the document, and Mrs. Kathy Berry assisted in the final formatting. Dr. Brice Stone and Dr. Thomas R. Saving provided many technical insights during discussions of the material.

NEURAL NETWORKS AND THEIR APPLICATION TO AIR FORCE PERSONNEL MODELING

SUMMARY

This report evaluates the potential for applying neural networks to Air Force personnel analysis and pinpoints those areas of personnel analysis most suitable for examination with neural network techniques. Neural network technology has recently demonstrated capabilities in areas important to personnel research such as statistical analysis, decision modeling, control, and forecasting. An extensive review of the neural network literature indicates that these networks have proven superior to more traditional analytic techniques in many applications. This review also indicates that three different neural network architectures are particularly suited to modeling many aspects of the Air Force personnel system. As demonstrated in the literature, the principal benefit offered by these architectures is the ability to derive nonlinear and interacting relationships among the components of a model. The three networks described in the report (back propagation, learning vector quantization, and probabilistic neural network) are all shown to be capable of representing much richer relationships than those obtained by standard parametric models.

Combined with an examination of current Air Force personnel models, the review of neural network literature indicates several personnel modeling areas which could benefit from the added flexibility of the neural network architectures. In particular, two areas were selected to empirically evaluate the application of neural networks in personnel research: airman reenlistment decisions and airman inventory modeling. Conceptual models based on prior research in these areas were developed and the method of applying neural networks to these models is outlined in the report.

INTRODUCTION

This is the final report of a task to evaluate the potential of applying neural network technology to Air Force Personnel modeling. The nature of this task requires that this report address several rather disparate areas. The report serves as both an introduction to neural networks and a research plan for applying neural networks to the personnel system. Incorporated into this framework is a description of three important network architectures, along with a brief review of armed forces personnel models.

Recent non-military research in neural networks strongly suggests this new technology will have implications in several areas related to personnel planning and management. Neural networks have been compared against traditional techniques in several areas such as curve fitting and system control and found to surpass the capabilities of those techniques in many cases. Despite this extensive ongoing research in neural networks, no efforts are currently focused on manpower and personnel issues. One of the major goals of this task is to identify those areas in the Air Force personnel system which are most amenable to the application of neural network techniques and to suggest areas where neural networks can be effectively compared with more traditional methods. A secondary goal involves the introduction and explanation of neural network techniques to researchers and analysts in the Air Force personnel field.

During this research five major objectives were accomplished:

1. Survey and review of neural network techniques, methodology, and applications.
2. Review of Air Force personnel modeling.

3. Conceptual development of personnel system models using neural networks.
4. Identification and description of existing models or traditional methods against which neural networks can be compared.
5. Development of a primer on neural networks.

The field of neural networks is highly interdisciplinary and marked by great diversity in its models, techniques, and research goals. Some of the most successful techniques are described in Section II, along with a brief introduction to the general concepts of neural networks. Some current personnel models are reviewed in Section III, and particular attention is paid to areas where neural networks may prove useful. These models range in complexity from simple linear reenlistment functions to multifaceted simulations of the entire personnel inventory. Drawing on the information in the previous sections, several specific Air Force personnel models appropriate for examination with neural networks are outlined in Section IV. Plans for implementing the models using neural networks are discussed, and data requirements are outlined. Methods of evaluating and validating the resulting models have been previously documented in Stone, Looper, and McGarrity (1990). Several specific applications of neural networks are surveyed in a separate literature review (Wiggins, 1990a). The survey focuses on applications that are related to, and provide background for, potential applications in the personnel arena. In addition, Wiggins, Looper, and Engquist (1990) provide an introductory tutorial on neural networks.

INTRODUCTION TO NEURAL NETWORKS

Neural networks have a history dating from the turn of the century. However, their application, outside of physiological and some psychological research, was limited until the 1940s and did not begin in earnest until the 1980s. The driving force behind much of the neural network research has been the capability of the brain and nervous system to perform complex pattern recognition, control, and cognitive tasks.¹ Emulation of the highly distributed and interconnected nature of the brain may produce automata with some of the capabilities of biological neural networks. The networks of concern here are implemented as software or hardware simulations which are loosely based on our knowledge of the characteristics of biological neurons. These networks are often referred to as artificial neural networks or ANNs to distinguish them from their biological counterparts. Although neural networks have been applied in areas ranging from associative memory to optimization and control, the focus in the present report will be on the general areas of classification, prediction, and control.

Three features or characteristics differentiate neural networks (both biological and artificial) from most other methods. First, neural networks are composed of simple processing elements. Second, many processing elements are employed to perform any task. Third, all of the elements process and communicate information at the same time. Taken together, the last two features define a distributed parallel computing system. This type of system is being explored in several areas such as parallel supercomputers. It is the use of a vast number of simple processing elements, an extremely high degree of parallelism, and automated learning methods which distinguishes neural networks from these other distributed parallel systems.

¹For a detailed survey of the historical development of artificial neural networks and early neurological research, see the collection of papers annotated by Anderson and Rosenfeld (1988).

Within these boundaries, there are many neural network architectures (or types of neural networks). Of primary interest are those architectures which allow the network to capture information from potentially noisy inputs and then, given new inputs which may represent novel situations, generalize their response using the information previously captured. A few of the specific areas where this capability has been exploited include: hand-written character recognition, stock price forecasting, classification of sonar signals, and control of robotic devices.

Artificial Neurons

The processing elements or neurons² which form a neural network are usually modeled as simple nonlinear functions. They accept a set of N inputs, compute the products of the inputs with a set of N weights, and pass the result through a nonlinear function referred to as a "transfer function." Figure 1 depicts a neuron that operates in this fashion. In this case, the neuron is operating on inputs which could be taken to represent important factors in an airmen's reenlistment decision. Each of the inputs (length of service, dependents, etc.) is multiplied by its associated weight, and a sum S is produced. This sum is then passed through a nonlinear transfer function. Three possible nonlinearities are shown: hard-limiting, sigmoidal, and threshold. The inputs could be different for another problem; or, in many cases, would be the outputs of other neurons instead of direct connections to the "outside world." Some neural network architectures postulate more complex neural functions: using spike trains rather than real numbers, accounting for temporal features, or employing more complicated aggregation functions than a simple weighted sum. However, the majority of current networks employ the "sum and fire" type of neuron shown in Figure 1. Most network architectures are differentiated by how the neurons are connected (network topology) and the rules for training or adapting the network to incoming signals or inputs.

Neural Network Architectures

There are over twenty major types of neural network architectures currently in use. Many of these major types also have several variations on their basic scheme. Specific architectures are usually most useful in particular problem domains: early vision, cognitive functions, associative memory, classification, function approximation, etc. A few have more general capabilities and applications. The first two architectures discussed below have proven to be some of the more useful in several different domains. They represent some of the most mature techniques in this very young field. In addition, their methods of capturing and representing information lie at opposite ends of the neural network learning spectrum. The third architecture, Probabilistic Neural Network (PNN), is particularly suited to classification problems and is based on established Bayesian classification techniques. These three architectures and their variants are prime candidates for application to personnel issues.

Back Propagation

One of the most widely applied neural networks is the back propagation³ architecture discovered independently by Werbos (1974) and Rumelhart, Hinton, and Williams (1984). This

² The simple processing elements that form a neural network are referred to using several different terms: processing elements (PEs), neurons, or computational elements.

³ A testimony to the relative youth of the neural network field and the variety of disciplines contributing to the field is the equivocation in the use of terms and even spellings. The most studied and applied architecture in the field will be found as back propagation, backpropagation, or back-propagation, depending on the author. Back propagation will be used in this report except in direct quotes and bibliographic references where the author's spelling will be retained.

architecture allows a network to learn complex nonlinear relations between its inputs and outputs by forming an internal representation in layers of neurons with nonlinear transfer functions. The term "back propagation" sometimes refers only to the method of learning described below; it is also often applied to the complete architecture of layered neurons operating in a feed-forward topology and trained by back propagation of errors. As can be seen in Wiggins (1990a), back propagation networks have demonstrated several capabilities—particularly in classification, control, and functional approximation problems. Given the prominent position of back propagation networks, they will be used to demonstrate many neural network concepts. Specific problems and potential solutions associated with these networks will also be treated in somewhat greater detail.

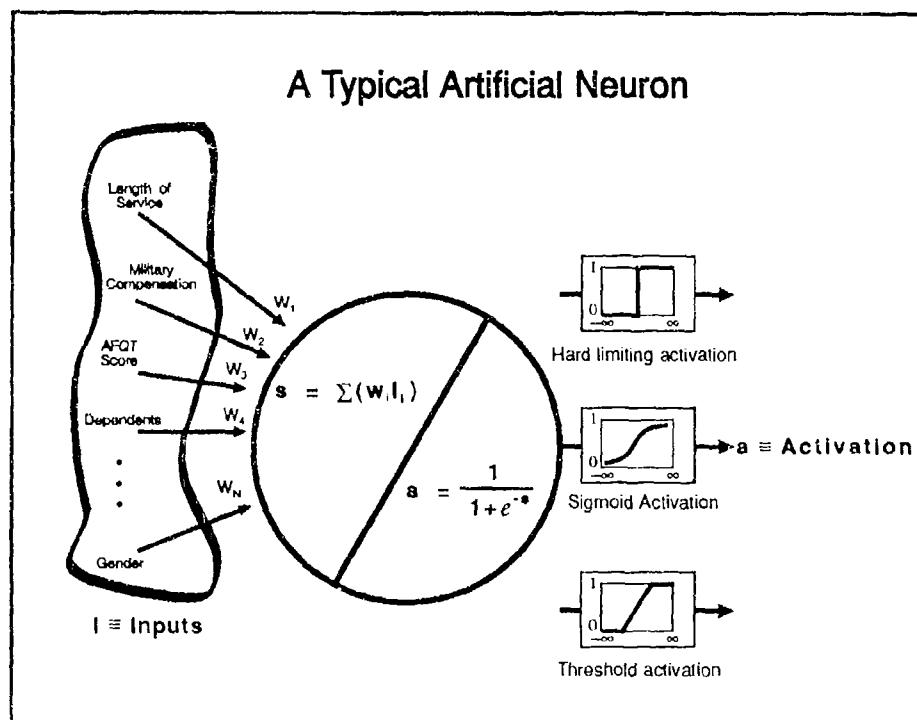


Figure 1. An artificial neuron with some reenlistment determinants as direct inputs. The neuron computes a weighted sum of the inputs and passes the result through a nonlinear transfer function. The forms of three alternate transfer functions are shown.

Back propagation is an error-correcting learning technique that seeks to minimize the prediction error of a neural network. This error is usually defined as the sum of squared errors (SSE) over all training exemplars.⁴ Other back propagation formulations are possible, such as maximizing likelihood or minimizing the absolute value of the errors (see Lippman,

⁴A training exemplar is a single observation of inputs and outputs to which a network is to be trained. It is directly analogous to observations or cases in regression analysis. Another term frequently used for exemplars is "training patterns." Although these terms are all interchangeable with respect to network operation, each usually has its own meaning in a particular problem domain.

1987). Minimizing the SSE is also the goal of most regression techniques; but, in the case of neural networks, the flexibility of the network allows more general models to be captured. Back propagation networks generally take on the form of the layered network shown in Figure 2. To facilitate the discussion, an example from the personnel system has been chosen for demonstration. An extremely simple airman reenlistment classification problem, using only two determinants (length of service and number of dependents), is shown. In addition, the size of the network is kept very small so the problem can be addressed without resort to vector notation.

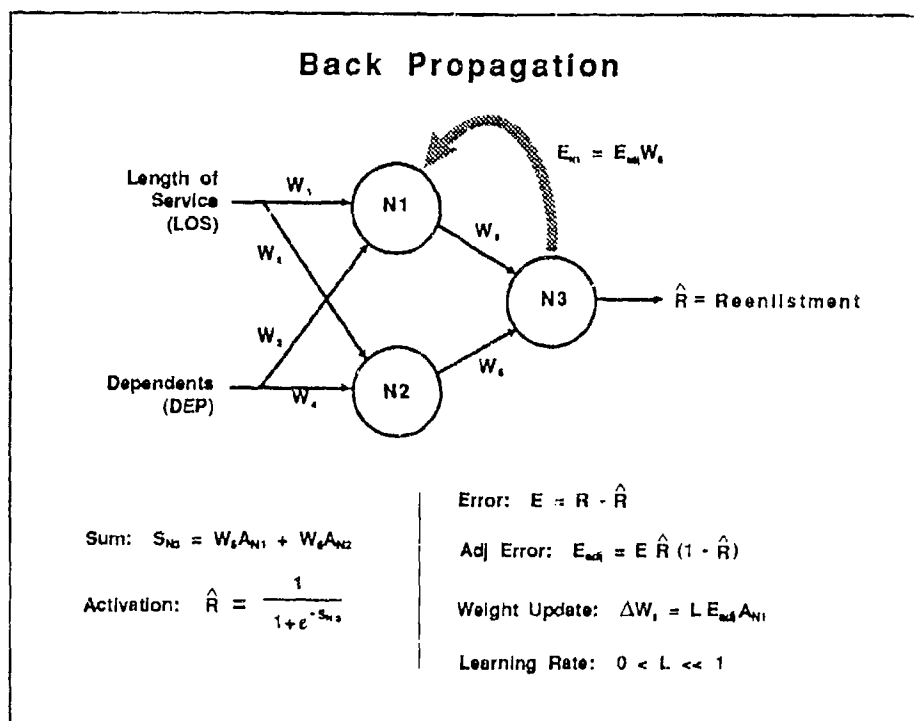


Figure 2. A simple back propagation network to predict reenlistment/separation decisions of enlisted airmen. The feed-forward equations are shown on the left, and the equations for weight adaptation are shown on the right.

The two neurons labeled **N1** and **N2** form a "hidden" layer which receives its signals (length of service and dependents) from the input layer. The neurons in the hidden layer pass their outputs (A_{N1} and A_{N2}) to the output neuron. The output layer in this case is composed of a single neuron **N3**. This output neuron computes its output based on these outputs from **N1** and **N2** and the connecting weights W_5 and W_6 . This flow of information from input to output is referred to as the "feed-forward process," and this type of network is called a "feed-forward architecture." Alternatively, networks that contain feedback connections from the hidden layers or outputs to prior layers are called "recurrent networks." It should be pointed out that the architecture of the network in Figure 2 is particularly simple. Typically there are more than two inputs, and often more hidden layers of neurons are employed. Each hidden layer's neurons are usually completely connected to the neurons in the previous layer (closer to the input). In addition, the output need not be limited to a single neuron. In the current example, if one wished to model the extension decision along with the reenlistment decision, two additional neurons (one for separation and one for extension) could be added to the output layer.

Training With Back Propagation

Training in neural networks is normally an adaptive process, with the network adjusting itself each time it receives a new exemplar. An illustration of this learning using Figure 2 will provide some insight into the process. For the example, one should assume training exemplars (observations) are available on individual airmen at a reenlistment decision point. Also, the observations include the airman's length of service, number of dependents, and reenlistment outcome (0 if the airman separates, 1 if he reenlists). An airman's length of service and number of dependents are provided as inputs to the network. The neurons **N1** and **N2** process the inputs by multiplying each input by its respective weight (W_1 and W_3 for **N1**, and W_2 and W_4 for **N2**). The resulting sums are passed through a sigmoid activation function to produce the output for each neuron in the hidden layer. These neurons are operating in exactly the same manner shown in Figure 1 when the sigmoid transfer function is used. These outputs are then fed into the output neuron **N3**, which performs the same summing and transformation function. These functions are shown for **N3** in the **Sum** and **Activation** equations in Figure 2. After this feed-forward process, the output of **N3** is interpreted as the classification prediction for the airman. The output is a real value in the range of 0 to 1. If the output is above .5, the network is predicting a reenlistment; if it is below .5, the network is predicting a separation.

During the training stage, the network is also provided with the actual decision of the airman. Given this actual decision and the predicted decision of the network, the back propagation training algorithm attempts to adjust the network so that its response is closer to the airman's observed decision. Toward this end, **N3** computes its output error E as shown in Figure 2. This error is adjusted by the derivative of the activation function, and the adjusted error is used to adapt each of the neuron's weights by a small amount, determined by the learning rate. This adjustment causes the neuron's output to be closer to the observed decision of the airman. Thus far only the weights on the output neuron have been adjusted. Because the neurons in the hidden layer do not have a target output, it is not initially clear how to adjust their weights. This is known as the credit assignment problem. Back propagation employs the chain rule of integral calculus to assign some of the blame for the final output error to the hidden neurons. As seen in the figure, **N1** is assigned an error E_{N1} proportional to its contribution to the final output. This error can then be used by **N1** to adjust its weights using precisely the same process used by the output neuron. Likewise, **N2** follows the same process. The learning rate L determines how much adjustment is made by the neurons, and thus, how quickly they adapt to each new exemplar.

If the learning rate is small enough, the algorithm outlined above implements a first-order gradient descent search in weight space for the set of weights that will minimize the sum of squared errors over the outputs for all exemplars in the data set. In other words, the algorithm seeks that set of weights which produces the closest fit to the observed decisions using least squared error as the fit criterion. The training process can be slow and, in the case of difficult problems, can require several thousand passes through the complete set of exemplars before the weights stabilize and the network converges.

More formally, the SSE criterion can be expressed as:

$$E = \frac{1}{2} \sum_o \sum_j (te_j - o_{oj})^2 \quad (1)$$

Where:

- t_{ej} is the target or desired output of output neuron j for the exemplar e .
- o_{ej} is the output of neuron j (neuron j in the output layer) for exemplar e .
- E is the total error across all output neurons and all training exemplars.

Gradient descent requires that each weight change be proportional to the impact of the weight change on total error E ; thus:

$$\Delta w_{ji} \propto - \frac{\partial E}{\partial w_{ji}} \quad (2)$$

Where:

- w_{ji} is the weight from neuron i to neuron j .

Because the output of a neuron for a given exemplar is merely:

Where:

$$o_{ej} = \frac{1}{1 + e^{-\sum_j w_{ji} o_{ei}}} \quad (3)$$

- o_{ei} is the output of neuron i in the layer feeding into the layer containing neuron j (this may be a hidden layer or in some cases a direct input), differentiating Equation 1 with respect to o_{ej} and Equation 3 with respect to w_{ji} , then combining the result with the chain rule produces

$$\frac{\partial E}{\partial w_{ji}} = \sum_e (t_{ej} - o_{ej}) o_{ej} (1 - o_{ej}) o_{ei} \quad (4)$$

This is precisely the value required for application of gradient descent as shown in Equation 2. This derivative requires the observed target value for each exemplar t_{ej} and is applicable only to neurons in the output layer. The first component in the summation is simply the prediction error of the output neuron, whereas the second component is the derivative of the sigmoid activation function. It can be seen that the weight adaptation rule shown in Figure 2 performs precisely the update required in Equation 4 (with the learning rate as the constant of proportionality). Obtaining the derivative for neurons in the hidden layer requires another application of the chain rule and produces an expression which requires the back propagation of errors shown by the large arrow in Figure 2.

The derivation above assumes the network is presented with all of the exemplars before the network's weights are updated. This process is known as "batch learning." Adapting the weights after each exemplar, as shown in Figure 2, is referred to as "on-line learning." The learning rate would have to be infinitely small for on-line updating to follow the actual gradient from all the exemplars. Conversely, on-line learning is continually estimating a local gradient based on the current exemplar. Rumelhart et al. (1984) present an informal derivation of back

propagation using on-line learning and also show a detailed derivation for the adaptation of hidden neurons. Both forms of learning are used in practice, and neither has proven consistently preferable in all cases.

The final result of this derivation is an algorithm (shown in Figure 2) for performing gradient descent in a layered network using only local information. It solves the credit assignment problem for neurons in the hidden layers, which allows the learning of nonlinear functions. Creative application of the chain rule, and the use of simple gradient descent, allows each neuron to adapt its weights using only information from the neurons to which it is directly connected. By freeing the network from the need for global information, the back propagation algorithm allows the network to be implemented in parallel using a very fine grain size-by-neuron.

Capabilities of Back Propagation

The example above used back propagation to classify airmen according to their expected reenlistment/separation intentions. In practical applications, the continuous output of the final neuron is usually interpreted to represent the confidence of the classification or the probability that the positive result will occur (airman reenlists). In these types of classification tasks, a feed-forward network with two hidden layers can produce an arbitrarily complex decision region to classify the inputs. The region can contain non-convex partitions, and individual classes can form discontinuous partitions.⁵

If the sigmoid transfer function on the output neuron is changed to a linear function, the network can produce real valued results spanning the real number system (Lapedes & Farber, 1987). This architecture allows the network to model any system that requires a mapping of inputs to outputs. In fact, several researchers have shown that a feed-forward network with at least one hidden layer, and monotonically increasing nonlinear transfer functions, can produce any continuous mapping of inputs to outputs (Funahashi, 1989; Hecht-Nielsen, 1987c; Hornik, Stinchcombe, & White, 1989). This is probably one of the most important theoretical results in the field. It demonstrates that feed-forward neural networks can be used as universal function approximators. Any continuous functional form can be captured and reproduced by the interconnections in such a network.

This result holds particular promise for problem domains where the inputs to a system (or decision) are known, but it is impossible to theoretically determine the form of the relationship between the inputs and outputs. The personnel system is rife with such examples. How does the unemployment rate affect an airman's decision to reenlist? Does gender affect the impact military compensation has on a potential recruit's likelihood to enlist? In fact, it is almost impossible to find a case where the functional relationship (linear, log-linear, exponential, etc.) is known. It is even more difficult to specify whether the determinants' effects are interrelated (e.g., an airman may be sensitive to civilian wages only when the unemployment rate is sufficiently low). The ability of a feed-forward neural network to produce any required relationship that fits the observed behavior of a system could be very important in these areas. The form of the model itself becomes data-driven rather than simply representing the parameters of a predefined functional system.

⁵Some examples of these types of regions are discussed in Wiggins and Looer (1990), a neural network tutorial.

Back Propagation Problems and Solutions

Local Minima. In the form described above, back propagation has several theoretical and operational problems. McInerney, Haines, Biafore, and Hecht-Nielsen (1989) have demonstrated that back propagation networks can exhibit local minima in their error surfaces. This has significant implications for the convergence of the algorithm. A feed-forward network can be a universal approximator; however, under conditions with local minima, the back propagation training algorithm is not guaranteed to find the best approximation for a given network structure.

Avoiding Local Minima. There are no solutions to the problem of local minima while remaining strictly within the framework of gradient descent search used by back propagation. Any gradient-following system, whether first- or second-order, is subject to becoming trapped in local minima (if such minima exist). Rumelhart and McClelland (1986) claim that local minima are unlikely to occur in networks with many hidden units. The added degrees of freedom in such networks, by increasing the dimension of the search space, actually increase the likelihood that the search will be over a convex surface.

Baba (1989) has suggested the use of a random optimization method (Matyas, 1965) to avoid the problem of local minima. Baba's recommendation is to generate a set of Gaussian random errors and add those to the weights in the network. If the fit of the network improves, keep the change; otherwise, return the network to its original state. This is a straightforward random search technique and guarantees the convergence of the network to its global minimum error (Solis & Wets, 1981). In his empirical tests, Baba found that the algorithm performed faster and found the global minimum more reliably than did back propagation on two of three example problems. However, and particularly on one example problem, the speed and ultimate convergence of Baba's method were highly dependent on the choice of the variance of the Gaussian errors. Patrikar and Provence (1990) suggested a very similar technique which involves adding a random perturbation to a single weight and accepting the change if the network's performance improves.

Whitley and Starkweather (1990) have suggested the use of genetic algorithms to search for the weights in a feed-forward network. These algorithms operate by maintaining a population of solutions to the problem (weights in this case) and allowing these solutions to selectively exchange information based on which solutions are most "fit" for the problem (see Goldberg, 1989; Holland, 1975). Although they do not guarantee the global minimum, genetic algorithms are expressly designed to search error surfaces with many local minima and find "good" or near-optimal solutions. Early empirical results from Whitley and Starkweather are encouraging. It should be noted that neither of these solutions solves the problem that back propagation encounters when local minima are present. Rather, completely different search techniques are substituted for back propagation. Still, both techniques lend themselves to distributed hardware implementations and the vast increase in speed such architectures offer.

Slow Convergence. Related to the local minima problem is the very slow convergence and consequent long training times of the back propagation algorithm. It is not uncommon for back propagation to require 20,000 to 30,000 passes through a data set before the weights converge. This slow convergence often results from long, gently sloping regions in the error surface. These regions also make it difficult to determine when the algorithm has converged. Weights may remain very stable and little reduction in SSE may be observed over long training sequences as the algorithm passes over such a surface.

Speeding Convergence. Given the desirable properties of the algorithm, the problem of slow convergence has received extensive attention in the literature. It should be noted, however, that speed is a problem only when training back propagation networks. Once a network has been trained, computing the result, prediction, or classification for a new set of inputs is straightforward and rapid. Most of these speed ups take one of three forms:

1. Standard optimization techniques.
2. Heuristics for adapting network training parameters.
3. Order and selection of training exemplars.

The most pervasive suggestions for increasing convergence rates using the back propagation algorithm involve the use of optimization techniques. Back propagation employs one of the simplest of optimization techniques—first-order gradient descent. Most efficient optimization techniques utilize some second-order information about the gradient, and these are the most common suggestions for speeding up back propagation. Several researchers have suggested more traditional curve-fitting techniques which use second-order information: recursive least squares (Kollias & Anastassiou, 1988; Palmieri & Shaw, 1990) or Kalman filtering (Scalero & Tepedelenlioglu, 1990; Singhal & Wu, 1989). Though these techniques are often efficient, they require complete information on the entire weight matrix to update each weight.⁶ This requirement makes the techniques much more difficult to implement in parallel hardware (especially the fine-grain parallelism associated with neural networks). Less restrictive techniques have been suggested that estimate second-order effects using only local information. Kramer and Sangiovanni-Vicentelli (1989) and Cho and Kim (1990) have suggested various forms of conjugate gradient algorithms. The work of Fahlmann (1988), Becker and Le Cun (1988), and Dewan and Sontag (1990) can be best described as quasi-Newtonian methods. Line search algorithms have also been proposed (Dahl, 1987). These are only a handful of the hundred or so hybrid second-order techniques that have been explored. The empirical results from each of these techniques typically demonstrate significant speed improvements over standard back propagation. Five- to 50-fold increases in convergence speed are not uncommon using these methods on selected problems.

A second common method for accelerating the back propagation algorithm involves adapting the learning parameters. Most important among these parameters is the learning rate L shown in Figure 2. The convergence rate and stability of the network can depend dramatically on the value of this learning rate. The rate is usually set at a fixed value, or follows a simple declining schedule as learning progresses. When the rate is allowed to adapt to the local slope of the error surface, significant performance increases have been found. Several researchers have suggested heuristics for adapting a global network learning rate (Battiti, 1990; Cater, 1987; Chen & Mars, 1990; Vogl, Mangis, Rigler, Zink, & Alcon, 1988). In general, these heuristics take the form of rules which increase the learning rate when it appears the network is in a flat region of the error surface. Jacobs (1988) extended this line of research and developed heuristics for adapting a separate learning rate for each individual neuron. This method was subsequently refined by Minai and Williams (1990).

The third method often used to accelerate training involves selecting and ordering the training sample. Lippman (1987) carefully chose equal numbers of exemplars from each class in a classification problem. He also ordered the sample such that the classes alternated on each presentation of an exemplar. Hoskins (1989) suggests "focused-attention backpropagation," which selects for presentation exemplars that are difficult to learn. Essentially, the network ignores those exemplars which it can correctly classify and trains only on those it is currently misclassifying. Several variations on presentation order have been examined by Ohnishi, Okamoto, and Sugie (1990). Speed improvements over standard back propagation on sample problems ranged from none to threefold increases.

⁶A complete bordered Hessian matrix of weights must be inverted for each step toward the final solution.

Other methods to accelerate back propagation have been tried. Storretta and Huberman (1987) adjusted the sigmoid transfer function to be symmetric about zero. Along these same lines, Rezgui and Tepedelenlioglu (1990) used a limited-range linear activation function. Unlearning (or weight decay) during training was suggested by Hagiwara (1990). Baba (1990) combined his random optimization method with gradient descent to speed convergence. Samad (1990) viewed the back propagation algorithm as a series of rules and suggested several logical variations on those rules.

Each of these techniques has demonstrated speed improvements over standard back propagation on some example problems. The speed increase sometimes reaches a factor of 50. However, there are usually problems for which the same methods demonstrate little or no improvement in speed; and some of the methods occasionally exhibit pathological behavior (wild oscillations or inability to converge). Widrow (1990) has pointed out that, though sometimes slow, gradient descent is an extremely robust technique when applied to convex optimization. These speed enhancement techniques may prove important if large quantities of data from the personnel system are to be routinely analyzed using only software simulations of neural networks.⁷ For the current research, the question of whether to use these techniques (and which techniques to use) is less pressing than assessing the applicability of back propagation itself.

Poor Generalization. Another area of difficulty involves generalization, or the ability of the network to perform well on exemplars not in its training set. This is a problem only if the underlying model is stochastic or there is noise in the data set. Many of the current applications involve engineering-type problems where there is little noise in the inputs and the model does not contain a large stochastic element. In this case, a model that fits the known data generally performs well on new examples within the range of the training data. Personnel problems, on the other hand, usually involve a substantial stochastic element. On problems with similar "noisy" elements, Rumelhart and McClelland, (1986) have found cases of back propagation fitting the training data well, but performing poorly on new exemplars. Preliminary analysis of individual airman reenlistment decisions and pilot Undergraduate Pilot Training (UPT) success has demonstrated distinct generalization problems (Wiggins, 1990b). Because of their flexibility, feed-forward networks can be prone to overtraining in these cases. Essentially, the network can "memorize" a data set, including the noise in the observations. The inclusion of this noise in the network's internal model degrades its ability to perform outside the training sample. The problem is related to overfitting in other estimation techniques.

Improving Generalization. This remains one of the least-addressed aspects of back propagation learning. Most early proposals to address the problem involved using small networks with a minimal number of neurons in hidden layers (again see Rumelhart & McClelland, 1986). A network with few neurons has less flexibility and therefore can learn only the main statistical features in the data set. Because the main features are exhibited by most of the exemplars and the noise or stochastic factors vary across exemplars, the smaller network is forced to ignore small differences in exemplars and is more likely to learn the characteristics of the "true" model. It is very common to try several networks with differing numbers of hidden layers and neurons in those hidden layers. Though less arduous, this behavior bears a strong resemblance to performing a specification search when doing regression analysis.

⁷Currently most research is done with software neural network simulators. Reasonably priced hardware will soon be available to implement some network architectures directly. These will run at 1,000 to 1,000,000 times the speed of software simulation and render the 5- to 50-fold speed improvements of these techniques less valuable for most problems.

In the same vein but removing the selection of network size from the researcher, Kruschke (1988) suggests several metrics for dynamically disabling specific nodes and weights during training. His methods attempt either to excise redundant neurons or to compress the dimensionality of the hidden layer. A more complicated method has been recommended by Mozer and Smolenski (1989). They specify a relevance metric which computes the impact of removing a neuron or weight on the error function for the network. Neurons or weights with little impact are removed during training. Other researchers have made similar suggestions (Bailey, 1990; Sietsma & Dow, 1988). All of these methods start with large, highly flexible networks and dynamically prune away redundant or unimportant neurons or weights. In all cases, the size of the resulting network will still depend to some extent on the setting of parameters that determine how thorough the pruning will be. Ash (1989) has developed an algorithm that proceeds in the opposite direction. He starts with the smallest network and adds nodes until the problem is sufficiently solved. To recognize a sufficiently solved condition requires the use of a holdout or test sample which is not included during training.

A very different method has been proposed by Lincoln and Skrzypek (1990). They tested the use of multiple small back propagation networks operating simultaneously on the same problem. On an abstract test problem, the multiple network model performed much better on unseen exemplars than did a single large network. Along different lines, Movellan (1990) examined the behavior of differing activation functions when three different noise distributions were added to equations representing missile ballistics. He found Tukey's distribution (Movellan, 1990) performed best and was much more resistant to noise than was the standard sigmoid activation function. He also found that exponential weight decay performed very similarly to Tukey's activation function.

Rumelhart (1990) has recommended several methods for improving generalization. The most theoretically based among these involves the addition of a weight penalty term to the error function (Equation 1). This method effectively enforces continuous decay of the weights in the network and is operationally similar to the exponential weight decay algorithm used by Movellan. Only those weights that consistently contribute to solving the problem will keep their values significantly different from zero. Several researchers have tested different specifications of the error function (Chauvin, 1990; Hanson & Pratt, 1989) and found that out-of-sample performance can be improved by this modification. Rumelhart's second suggestion involves maintaining a holdout sample. Training proceeds on the rest of the samples, and tests are performed at regular intervals against the holdout sample. When performance on the holdout sample degrades, training is stopped. Though simple, this method has proven empirically successful. Kimoto, Asakawa, Yoda, and Takewka (1990) employed the technique to predict stock market trends.

Along the same lines, Morgan and Bourlard (1990) examined the ability of networks of various sizes to generalize after varying amounts of training. They trained an array of networks ranging in size from 4 to 200 hidden units on two problems: a contrived classification problem with known noise characteristics, and a phoneme classification task using actual data. As training progressed from 1,000 to 10,000 training iterations, the performance of each network was tested on a holdout sample. The results indicated that both network size and amount of training had a significant influence on the ability of a network to generalize. They found that the out-of-sample performance of all the networks, regardless of size, degraded if training continued too long. Conversely, in-sample performance continued to improve with training. Smaller networks exhibited slower degradation and maintained a higher performance level even after extensive "overtraining." Still, over certain training ranges, the largest networks performed almost as well as the best-trained small networks. Morgan and Bourlard concluded that network size and amount of training should be determined empirically for each problem by maintaining a holdout or test sample for comparison purposes.

Early research indicates that the ability to generalize and techniques for obtaining good generalization will be critical in Air Force personnel applications. Unfortunately, this is an area with virtually no theoretical results and meager empirical support. The dynamics, and thus the training path, of back propagation learning are still not well understood. Despite these reservations, preliminary empirical work on using the techniques outlined above has produced encouraging results.

Learning Vector Quantization

Learning Vector Quantization (LVQ) is representative of a class of neural networks whose theory and implementation are quite different from those of back propagation networks. Where back propagation forms a global distributed representation of the inputs using all of its weights, LVQ forms local representations of the inputs in specific neurons. Kohonen (1989) developed the LVQ architecture to solve classification problems where cases or exemplars are to be selected into categories. Each exemplar is associated with the reference vector neuron whose weights are closest to its own inputs. The exemplar is then assumed to behave in the same manner as this reference vector neuron. This process is very similar to the nearest neighbor algorithm, which compares each new case to be classified with all known cases in the training data set. The new case is then assumed to fall in the same class as the closest case from the training data set (see Duda & Hart, 1973). LVQ can also be viewed as an extension of K-means clustering methods (Hartigan, 1975). K-means clustering has a goal that is similar to that of a version of LVQ: Find a set of reference means which partitions the input space such that intra-partition variance is minimized and inter-partition variance is maximized.

Competitive Learning in LVQ

The neurons in an LVQ network are adapted such that their weights become reference vectors which attract specific exemplars. The process can be described by referring to Figure 3. Again, a simple reenlistment decision example will be used. In this case, the airman's reenlistment military compensation (RMC) and the prevailing unemployment rate (UNEMP) are assumed to be the inputs or determinants for the classification. A very simple two-input model is used to facilitate a visual interpretation of the results. The architecture can handle an arbitrary number of inputs, and this extension is straightforward. This problem also has only two classes: reenlist and separate. This architecture is particularly well suited to problems with a large number of classes. As can be seen in the figure, the reference vector neurons are divided into two groups: those which classify reenlisters (the top group), and those which classify separators (the bottom group). The weights connecting these neurons to the inputs form the neuron's reference vector for the inputs. For example, the weights on the first neuron, W_{1R} and W_{1U} , are reference values, or attractors, for RMC and UNEMP, respectively. When an exemplar (an individual airman) is presented to the network, each neuron computes its distance from the exemplar. Euclidean distance, as shown in the calculation of the output for the sixth neuron (ON_6), is the most commonly used distance metric. The neurons then compete to claim the new exemplar, with the closest neuron winning the competition.

As Kohonen (1984) points out, it is possible to normalize the input vector to unit length. Once normalized, the distance calculation becomes a simple inner product computation with the weights. This makes the neuron's behavior just like that in Figure 1, with direct output of the sum (a transfer function is not needed). This pre-processing stage is left out of Figure 3 to simplify the discussion. The competition process itself can be implemented in parallel as a neural network, or a simple serial selection of the minimum can be performed (see Grossberg, 1973).

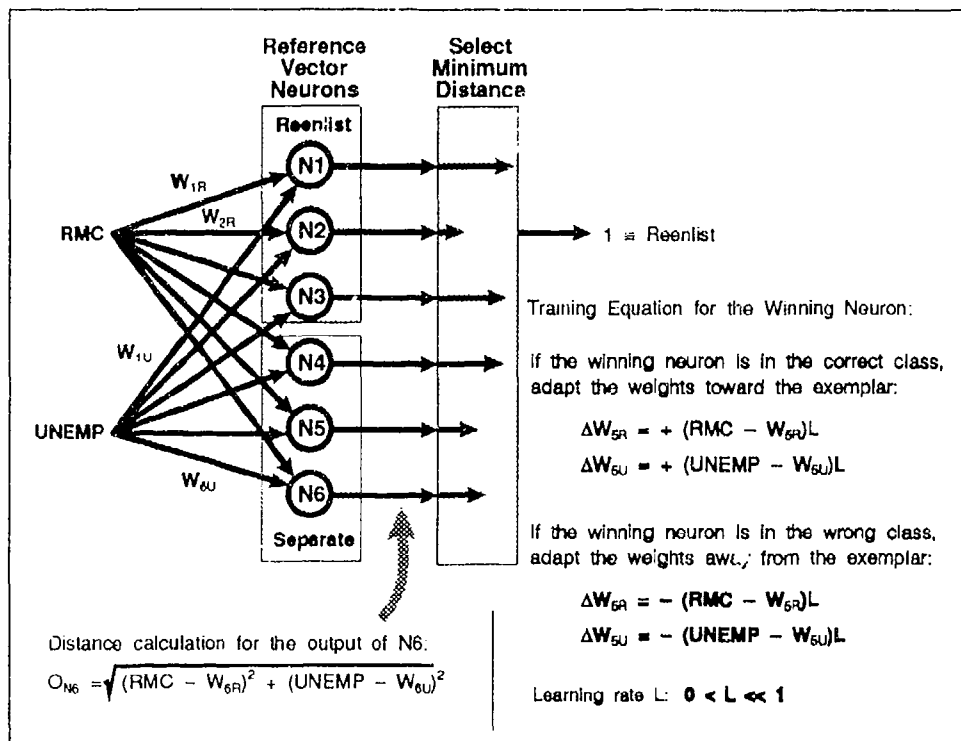


Figure 3. Schematic and computations for LVQ. Each reference neuron computes its distance from the exemplar, with the closest neuron adapting its weights toward or away from the exemplar as shown.

During training, the winning neuron adapts its weights toward or away from the input values of the captured exemplar. As with back propagation, the training is supervised and depends on the observed outcome (reenlistment/separation decision for the airman). If the winning neuron is a reenlistment neuron (from the top three in the figure) and the airman was observed to reenlist, then a correct classification has been made. In this case, the neuron adjusts its weights to be closer to the captured exemplar. As seen in the right of Figure 3, the adjustment is a simple linear proportion of the difference between the exemplar's inputs and the neuron's current weights. A small learning rate, which declines as training progresses, determines how far the weights are adjusted toward the exemplar's input values. If the neuron had misclassified the exemplar (a reenlist captured by a separate neuron, or a separator captured by a reenlist neuron), the neuron would adjust its weights away from the captured exemplar. In this method, the neurons move toward the centroids of regions where their classifications are correct and away from regions where their classifications are incorrect.

The effects of this training can be seen visually in Figure 4. A hypothetical distribution of airmen is shown. Each airman is marked by an S or an R representing separator and reenlist, respectively. In the top half of the figure, decision makers are shown distributed according to their military compensation and the unemployment rate at the time of their decision. The bottom half of the figure shows the final position of the reference vector neurons from Figure 3 after training. (The shaded area is the decision region for reenlisters.) As can be seen, the neurons form linear discrimination lines with their neighboring neurons. If there were four or more inputs, the discrimination surfaces would be hyper-planes. In this manner, piecewise linear decision regions are formed for each class. Because only six neurons were used in this example, the decision regions are very coarse. They can, however, be very flexible and even discontinuous if required by the particular problem.

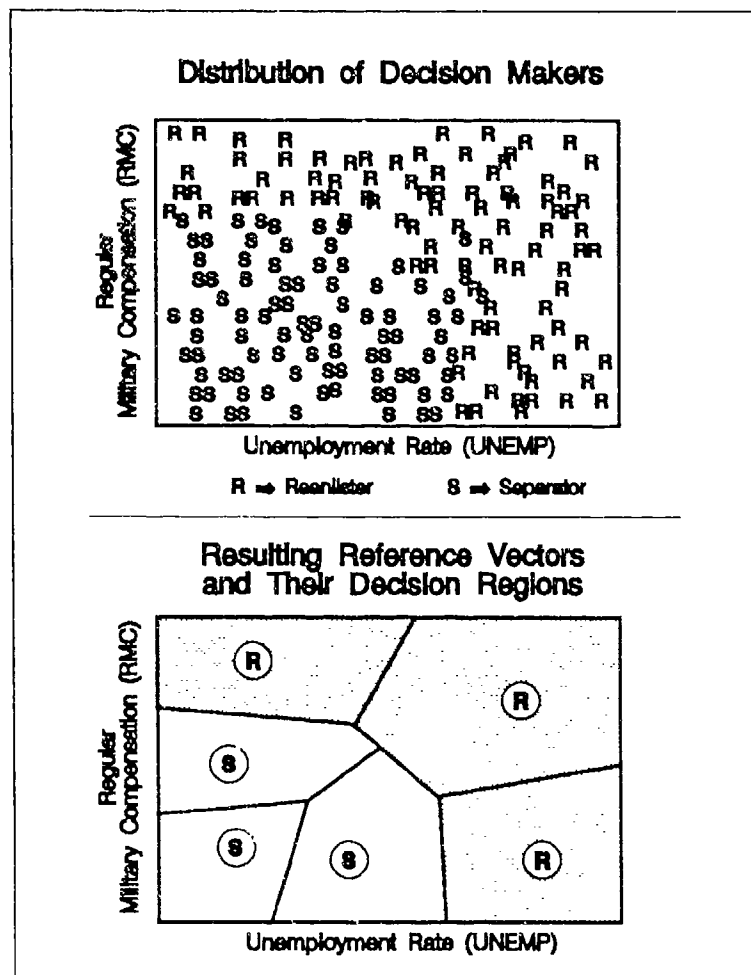


Figure 4. A hypothetical distribution of airmen at a reenlistment/separation decision point and the decision regions formed by applying the LVQ architecture of Figure 3 to this distribution.

Bart Kosko (1990) has used stochastic calculus to prove that a broad class of competitive learning algorithms converge exponentially quickly to the centroids of the inputs. LVQ is one of many algorithms which are subsumed by Kosko's derivation. The proof is similar to the application of Kolmogorov's theorem to feed-forward networks in that the centroids are defined to be only locally optimal. Even so, it guarantees stochastic convergence of the LVQ algorithm.

Related Architectures and Improvements

Related Architectures. LVQ is merely one example of a whole family of competitive learning neural network architectures. Unsupervised versions of the LVQ have been utilized to cluster exemplars without regard to known classifications (Kohonen, 1982b). Kohonen (1982a, 1984, 1989) has also developed an unsupervised version of the algorithm in which the neurons are arranged in a two-dimensional lattice. Neighboring neurons are adapted together, and the network forms topological feature maps similar to those for the cortical surface of mammalian

brains. This architecture has been particularly useful for developing internal representations of high-dimensional inputs.

Improvements. Kohonen (1990) has introduced several adjustments to improve convergence, class separation, and stability of the algorithm. Other researchers have also made similar suggestions (Darken & Moody, 1990; DeSieno, 1988; Kangas, Kohonen, Leaksones, Simula, & Venta, 1989). LVQ is also similar in spirit to the relatively new neural network architectures using receptive fields (see Moody & Darken, 1988). Rumelhart & McClelland (1986) examined competitive learning using somewhat different procedures and in various contexts. Though more biologically motivated, Grossberg (1973, 1986) has contributed many neurologically plausible, competitive architectures.

Hybrid Networks. Unsupervised versions of LVQ have been used in combination with other types of neural networks to produce several hybrid architectures. Hecht-Nielsen's counter-propagation network is the best known of these hybrids. The network is capable of producing arbitrary vector-to-vector mappings like the multilayer back propagation network. Hecht-Nielsen (1987a, 1987b) combined an unsupervised LVQ network with a Grossberg (1969, 1982) outstar network. In this context, the outstar operates in much the same manner as a simple, linear back propagation network. The network first uses the unsupervised LVQ to cluster the inputs into neighborhoods of related inputs. The outstar then learns a linear mapping from these neighborhoods for the desired output space. The nonlinearities of a problem are captured in the neighborhood clustering rather than the outstar weights. The counterpropagation network trains faster than the back propagation network but, in its normal configuration, is slightly less accurate for most problems. By contrast, de Bollivier, Galliari, and Thiria (1990) stack the networks in the reverse direction. They place a partially trained back propagation network in front of an LVQ network. The outputs from the hidden layer of the back propagation network are used as inputs for the LVQ network. These researchers developed a gradient descent algorithm for training the stacked network and show that it performs better on a wider range of problems than does either LVQ or back propagation alone. Their network also trains considerably faster than a back propagation network.

Probabilistic Neural Network

Overview

The Probabilistic Neural Network (PNN) was developed by Donald Specht (1988, 1990) specifically to solve classification problems. PNNs utilize classical Bayesian decision rules and local estimators for probability density functions (PDF) which are implemented within the context of a neural network. The algorithm shares some conceptual features with LVQ in that it estimates the multidimensional density function for a class using local information from the training sample. Instead of employing reference vectors to estimate the PDF, a PNN actually stores the inputs of each exemplar in a neuron. The multidimensional spatial location of these exemplars can then be used to construct a PDF for each category in a classification problem. Once the PDFs have been constructed, an observation whose category is not known can be classified by selecting the category with the highest point density at the location of the unknown observation's inputs. Specht has shown that the decision boundaries formed by the PNN asymptotically approach the Bayes optimal boundaries (i.e., those boundaries that minimize misclassification expected risk).

Training and Classification With a PNN

The Bayes decision rule employed in the PNN minimizes expected risk or cost associated with the classification. Using a two-class problem⁸ and continuing with a reenlistment/separation example, the decision rule can be specified as:⁹

$$\text{reenlist if: } h_r l_r f_r(X) > h_s l_s f_s(X) \quad (5)$$

$$\text{separate if: } h_r l_r f_r(X) < h_s l_s f_s(X)$$

Where:

- h_r is the a priori probability of reenlisting.¹⁰
- h_s is the a priori probability of separating.
- l_r is the cost or loss associated with classifying as a reenlistee an airman who separates.
- l_s is the cost or loss associated with classifying as a separator an airman who reenlists.
- $f_r(X)$ is the multidimensional PDF for reenlisters.
- $f_s(X)$ is the multidimensional PDF for separators.
- X is a vector of inputs representing the dimension of the PDF and with which the exemplar is to be classified (number of dependents, RMC, gender, etc.).

This rule classifies an exemplar into the class with the smallest expected risk or loss. The classification is based on known PDFs for each class, losses associated with misclassification, overall proportions in each class, and the vector of inputs for the individual exemplar. In most cases, the loss values or functions (l_r and l_s) are assumed to be equal, and they can be dropped from the equation. In terms of reenlistment/separation, dropping the loss functions requires the assumption that all misclassifications are equally costly.

The decision rule can be seen graphically in Figure 5. For exposition purposes, the PDFs are assumed to be univariate, with the only input being the civilian unemployment rate at the time of the reenlistment decision. The two PDFs shown have already been "scaled" by the a priori probability of a decision maker being in each class (h_r and h_s). In this manner, the area under both pseudo PDFs sums to 1.0. When the decision rule from Equation 5 is applied, the decision boundary is seen to be at the intersection of the two scaled distributions.

⁸ Extension to the multi-class case is straightforward.

⁹ The notation in this example is consistent with the back propagation and LVQ examples. It differs somewhat from that used in Specht (1990).

¹⁰ Operationally this probability is usually taken to be the proportion of reenlisters in the training sample. This proportion is simply the expected value of the probability of reenlisting based solely on the data in the training sample.

New or unknown airmen who face a decision when the unemployment rate is lower than that at the intersection would be classified as separators; those to the right of the intersection would be classified as reenlisters. If the density functions are correct for each class, any other decision rule would be nonoptimal and fail to minimize the number of misclassifications. The Bayes rule minimizes misclassification when the loss functions are equal, or "loss" when different losses are assumed or imposed on different types of misclassification. For example, misclassifying an eventual reenlistee as a separator may be more "costly" than misclassifying an eventual separator as a reenlistee.

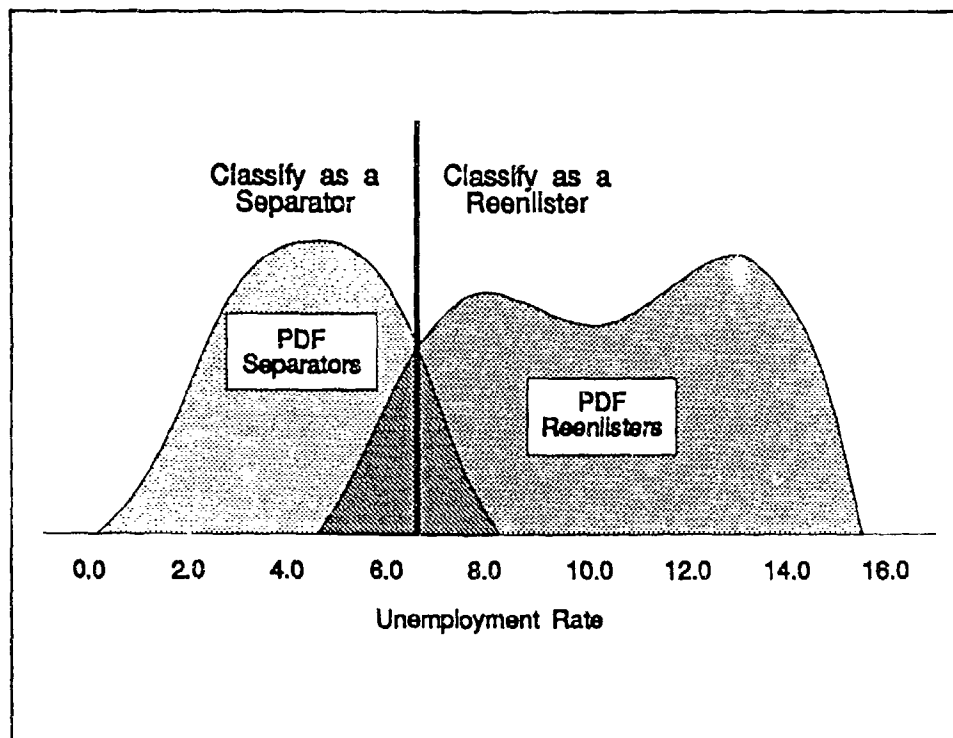


Figure 5. Application of the Bayesian minimum loss decision rule using hypothetical distributions of airmen at a reenlistment decision point. The boundary for classification of new or unknown decision makers is drawn at the intersection of the density functions for the two classes.

The decision rule outlined above can be easily applied if the PDFs for each category (reenlist and separate) are known. Estimation of these PDFs is analogous to training in the other networks and forms the core of the PNN. In PNNs, a PDF is estimated as the sum of many small multivariate Gaussian pseudo-distributions, each centered at a training example. Operationally, each training example in a class (say reenlisters) is stored, and the local density of the PDF is computed by measuring the distance from a new exemplar to all exemplars in the training set. The local density at any point on the PDF may be estimated as:

$$f_r(X) = \frac{1}{(2\pi)^{p/2} \sigma^p} \frac{1}{m} \sum_{a=1}^m \exp \left[-\frac{(X - X_{Re})^t (X - X_{Re})}{2\sigma^2} \right] \quad (6)$$

Where:

- p is the dimensionality of the input space (i.e., the number of inputs: RMC, unemployment, etc.).
- σ is a smoothing parameter, which determines the size or extent of the Gaussian around each training exemplar.
- N is the number of training exemplars or observations.
- X is a vector of inputs for the point at which the density is to be measured (or the vector for a new exemplar to be classified).
- X_{Re} is an input vector for the reenlister training exemplar e .
- t is a matrix transpose operator.

This computation forms the local density as a sum of small Gaussian pseudo-distributions around each known exemplar in the class (reenlisters in this case). Despite the use of Gaussians, the resulting PDF can assume any form dictated by the distribution of reenlisters along the input vector X . This distribution and the smoothing parameter σ dictate the final form of the PDF. The smoothing parameter determines the variance of the Gaussians or the effective range of each training exemplar. Specht (1990) has shown that as σ approaches infinity, the overall PDF approaches a multivariate Gaussian distribution. When σ approaches zero, any new exemplar is classified with its closest training exemplar. At this point, the PNN operates as a nearest neighbor classifier. The smoothing parameter effectively defines the size of the neighborhood around an unknown point, which will be used to determine the class of that point.

Figure 6 demonstrates the impact of changing the smoothing parameter while using the same five observations as a training sample. In this univariate example, five equally spaced observations are used to construct four different PDFs. Given the consistent training sample, the shape of the PDFs is determined solely by the value of the smoothing parameter σ . With a very small σ (the top PDF), the individual Gaussian kernels around each observation are apparent. As σ is increased, the impact of each observation becomes less localized and the total PDF becomes smoother. The value of σ is usually determined by empirically analyzing its effect on the performance of the PNN. Specht (1990) notes that classification performance of the PNN is fairly insensitive to changes in σ and fairly wide ranges of the parameter produce similar results.

The matrix multiplication in the numerator of the exponential function (Equation 6) actually serves to compute the squared distance of the new observation's input vector (X) from a given training exemplar (X_{Re}). As was the case for the LVQ network, this process can be reduced to a simple inner product between the new input vector and that of a training exemplar. Again, this is accomplished by normalizing all training and testing exemplars' input vectors to unit length. Once this is done, the estimation of the PDF can be easily performed in a feed-forward network where each neuron stores a training exemplar (see Specht, 1990).

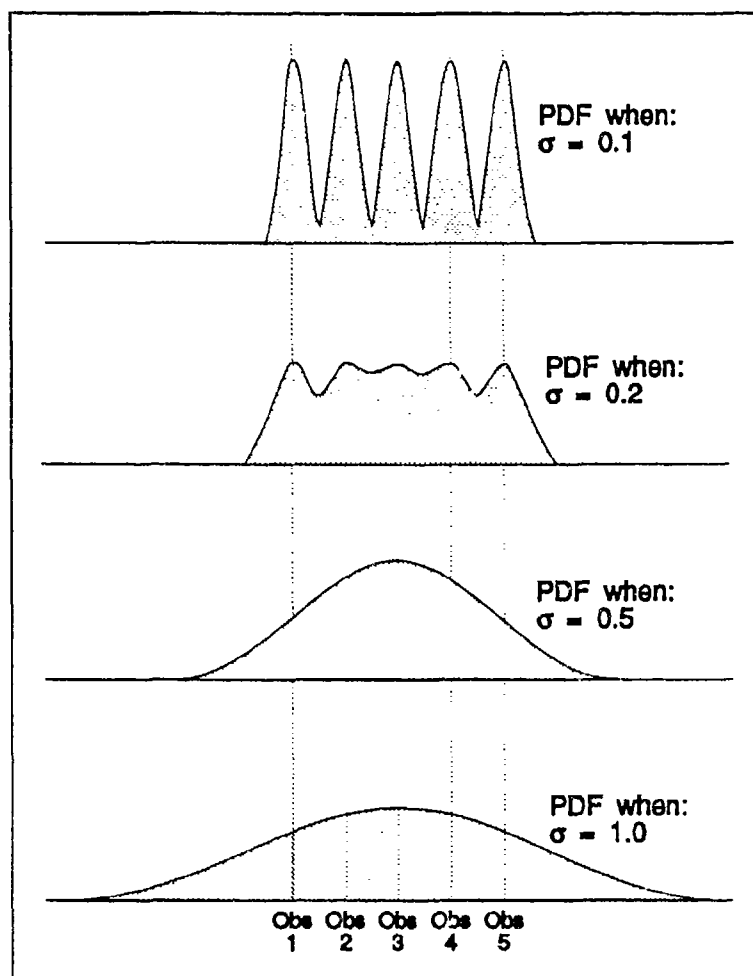


Figure 6. Effect of changing the smoothing parameter σ on the form of an estimated PDF. All four PDFs are derived from the same five sample observations.

Other Architectures

Many other neural network architectures have been developed. The three described above are some of the most generally applicable and well-studied architectures. In addition, they represent a broad spectrum of neural network concepts: local representation, global representation, self-organizing structures, and error-correcting learning. These three architectures and their variants have the potential to be applied to many personnel problems. Lippman (1987) has written an excellent review article that discusses several networks and their relation to pattern classification. Other major reviews have been prepared by Kohonen (1988), Grossberg (1988), and Carpenter (1989). Recently, two introductory neural network books have become available. Wasserman (1989) provides an introduction to nine major neural network architectures in his book *Neural computing, theory and practice*. Simpson (1990) addresses over 25 architectures, using a consistent notation, in his book *Artificial neural systems: Foundations, paradigms, applications, and implementations*. He assesses the capabilities of each architecture,

describes applications attempted with each architecture, and provides copious references. The report on a neural network study performed by the Defense Advanced Research Projects Agency contains an overview of the technology as of February 1988 (Darpa, 1988). Many early applications are discussed in that study. Finally, Anderson and Rosenfeld (1988) have compiled a collection of 45 seminal articles published in the field between 1890 and 1987.¹¹ Despite the recent vintage of most of these reviews and introductions, the extremely rapid advance of information in the field already makes them somewhat dated with respect to the most successful variants of the architectures, theoretical analyses, and empirical results. Still, each provides an overview of concepts and methods upon which most of the current adaptations and results are based.

AIR FORCE PERSONNEL MODELING

The personnel system in the U.S. Air Force comprises a large number of interacting components whose primary goal is to maintain mission readiness. Personnel managers and planners in each area (e.g., accessions, promotions, assignments) seek to optimize the levels and location of qualified personnel according to manning requirements for each system. At the same time, individual airmen make decisions within the system (e.g., separation, extension) based on their own preferences and well-being. All of these decisions are being made in a complex environment where actions in one area, such as Selective Reenlistment Bonus (SRB) policy, can impact decisions in another area, such as promotion. Figure 7 shows a highly schematic view of the airmen and information flows in the enlisted personnel system. The solid arrows represent personnel flows from one enlisted inventory cohort to another, whereas the shaded arrows represent information flow and information feedback.¹² At least one flow in the system is primarily driven by Air Force policy and management decisions: promotion. The other flows represent varying combinations of individual airman decision making and explicit control by personnel managers. Reassignment is primarily driven by management decisions, with varying amounts of airman input (depending on the programs in place at the time). Separation, reenlistment, and extension are currently determined wholly by individual airmen decisions. Still, these decisions are made in the context of current Air Force policies (SRB, military compensation, etc.), the composition of the force (availability of career job reservations, etc.), and economic conditions in the civilian labor force. Accession and retraining are driven by a combination of individual and personnel management decisions.

The explicit and implicit flows of information in the system are more complex than the physical flow of personnel. The education level, demographic factors, and aptitudes of those in the force (as well as those who are in the accession recruiting pools) form a context which constrains the implementation of policies and the attainment of manning goals. In turn, the effects of these very policies shape current and future characteristics of the personnel inventory. Congressional budget constraints must be balanced with manning requirements and the current and future force composition to produce policies that attempt to meet the manning requirements.¹³ All education, demographic, aptitude, economic, and policy conditions are eventually reflected in the personnel inventory and in the environment in which individual and management decisions are made.

¹¹ They have a second collection of articles forthcoming from the MIT Press.

¹² One could as easily define education, aptitudes, and demographic characteristics as forming dimensions of a cohort (in addition to grade, YOS, etc.), but they are treated in this view as information about the cohort.

¹³ This view of the system completely ignores the equally interesting task of translating general defense requirements and specific system readiness into manning requirements, a task with its own constraints and information (some shared with the system currently being discussed)

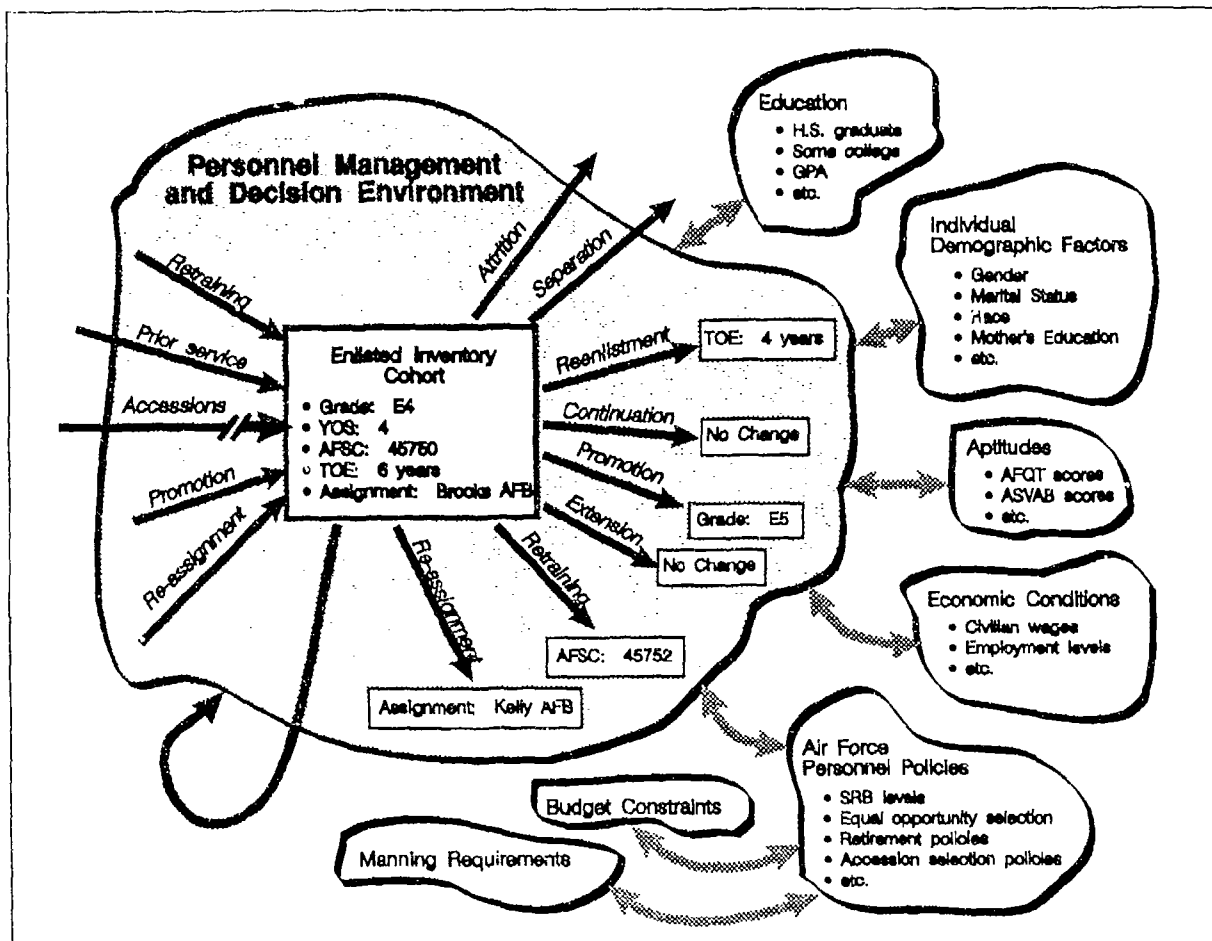


Figure 7. A conceptual view of the airmen and information flows in the enlisted personnel system. Solid arrows show the flows of airmen out-of and into a specific personnel cohort. Shaded arrows show the flow of information in the system and its feedback through implicit connections to all potential source and destination cohorts in the personnel system.

The job of modeling this system or its components involves abstracting the relevant features, dependencies, and interdependencies of the system or a subsystem from the complexity of the whole organization. The large number of factors affecting the personnel system, as well as the variety of individual decisions, management decisions, and policy decisions, make the personnel system extremely difficult to approach with any single modeling, simulation, or estimation technique. As with most complex systems, the personnel system is usually broken into smaller components for detailed analysis or aggregated to larger groups for analysis of the system as a whole. Decisions concerning which features to retain, which to ignore, and which to simplify determine the information content of a model. Explicit definition of both the retained features and the form of their relationships defines the conceptual structure of a model. This conceptual structure places bounds on the types of problems and levels of detail

for which the model is useful. Sometimes a conceptual model combined with theoretical derivations of its behavior is sufficient to address, at least partially, a particular problem. More often, specific quantified relationships between the components of the model must be found. In some cases, this quantified relationship is sufficient; in others, however, the dynamic behavior of groups or individuals operating under the specified relationships must also be quantified. It is in these last two areas, where it is important for models to capture relationships found in historical patterns, that neural networks are expected to be the most useful.

Types of Personnel Models

The types of models employed in Air Force personnel research encompass a broad spectrum of goals and techniques. In general, these models can be classified into three broad categories: analytic or descriptive models, planning models, and programming models. Analytic models are used to describe or analyze a particular functional area. They serve to increase understanding of an area by establishing relationships and constraints within the area. In establishing these relationships, analytic models seek to describe a particular functional area and quantify various aspects of the area. They typically focus on a specific individual decision (e.g., reenlist/separate), a particular inventory flow (e.g., accession), or a particular policy (e.g., Selective Reenlistment Bonus). Statistical and policy-capturing methods are usually employed in these models to determine factors affecting the decision, flow, or outcome. Analytic models are some of the most prevalent models in personnel research, and they have been applied to most parts of the personnel system. The process of extracting and quantifying salient features from a system increases understanding of the system and is a prerequisite to developing the two other types of models (planning and programming).

Planning models usually simulate the entire force, or some portion of the force, over time to assess the impact of policy or economic changes. Programming models are typically employed to determine the specific allocation of personnel resources. Often the major difference between a programming model and a planning model is the temporal horizon. Most planning models extend at least to the end of the current Program Objective Memorandum (POM) cycle, and some analyze impacts as far as 30 years out. Conversely, programming models usually restrict their horizon to the remainder of the current fiscal year. In addition, programming models usually handle the force, or a portion of the force, at a much more detailed level than a planning model addressing the same areas. The distinction between the analytic models and the planning and programming models is also somewhat hazy. Most planning and programming models explicitly or implicitly include information from one or several analytic models. Currently, neural networks will be most useful in developing analytic models. In these areas, their ability to abstract complex relations from observed behaviors or actions can be best exploited. The resulting analytic models may then serve as the basis for richer planning and programming models.

Accession/Enlistment Models

Aggregate Accessions

The importance of recruiting and enlistment to all of the armed forces is displayed by the number of models developed to explain and predict behavior in this domain. Ash, Udis, and McNown (1983) analyzed aggregate accessions in each of the four branches of the service. They estimated 15 race-specific equations among the four services and aggregated Department of Defense (DOD) accessions using two-stage least squares based on data from 1967 to 1976. After extensive testing, Ash et al. found that the models tended to perform rather poorly outside the estimation sample. DeV. ly, Saving, and Shughart (1978) also estimated a series of

aggregate Air Force enlistment rate models. These researchers included many additional factors not considered by Ash et al. and estimated models using both ordinary least squares (OLS) and grouped logit techniques. Their accession/retention model was later extended by DeVany and Saving (1982) to include endogenous recruit quality (measured by the Armed Forces Qualification Test) and waiting time effects (time spent in the Delayed Enlistment Program). Siegel and Borack (1981) examined an econometric model of aggregate naval enlistments, and Borack (1984) addressed the integration of supply models. In addition, documents and reports on several other models appear in Cirie, Miller, and Sinaiko (1981).

Applying Neural Networks to Aggregate Accessions

The Ash et al. research presents a model that is directly addressable by the back propagation network architecture. The independent variables employed by these researchers would serve as the inputs to the model. The known enlistment rates from 1967 to 1976 would serve as the training targets. In place of least squares estimation, a feed-forward network would be trained using back propagation. It is very likely that the "universal approximation" capability of back propagation networks would be important in this application. There is no theoretical or common sense reason the independent variables should have a strictly linear and independent impact on aggregate enlistment rates.

Given the relatively small data set, it is also very likely that some of the techniques to improve back propagation's out-of-sample performance would be required. Without these techniques, the flexible network architecture would tend to overfit the training data, to the detriment of the model's generalization performance. This method of applying neural networks directly in place of standard criterion-based estimators will hold for most of the personnel models to be discussed. The continuous nature of both the inputs (independent variables) and the output (dependent variable) of the Ash et al. model makes back propagation a natural network choice. However, variants of the PNN and LVQ techniques exist which can address this continuous vector-to-value mapping, and these techniques should not be dismissed out-of-hand.

Like the Ash et al. model, the DeVany et al. enlistment models could be directly "estimated" using the more flexible neural network methods. With regard to the simultaneity of some inputs, at least three possible approaches could be taken. The first-stage estimates of the endogenous variables could be obtained using OLS, as they are in two-stage least squares. These estimates have "removed" the endogenous effects and could be used directly as inputs to a neural network. Alternately, the first-stage estimates could themselves be formed from a neural network. The most likely solution would be to include all exogenous variables as inputs, including those used as instruments. The endogenous, "right-hand-side" variables would become the target outputs for the network. This process would effectively "estimate" a reduced form model.

Individual Enlistment

The enlistment behavior of high school seniors and recent graduates was analyzed by Hosek and Petersen (1986) using individual information from the 1979 DOD Survey of Personnel Entering Military Service and the 1979 National Longitudinal Survey of Labor Force Behavior (NLS). Hosek and Petersen estimated the DOD-wide probability of reenlistment using logit analysis based on survey responses. In a related study, Orvis, Gahart, and Hosek (1989) compared similar individual-based models to a regional cluster-based model. In general, the researchers found that the cluster-based models added little information to the models estimated on individual data.

Disaggregate DOD-wide enlistment has been considered by other researchers (Borack, 1984; Curtis, Borack, & Wax, 1987; Orvis & Gahart, 1985, 1989). Some of the Navy's experience with aggregate and disaggregate accession models is summarized in Cirie, Miller, and Sinaiko (1981), with further research documented in Cowin, O'Connor, Sage, and Johnson (1980). In addition, Verdugo and Berliant (1989) examined prime recruiting markets for the Army.

Applying Neural Networks to Individual Enlistment

The individual enlistment problem described above is a typical example of a classification problem. Each potential enlistee is to be classified as either a likely enlister or a likely non-enlister based on a set of individual characteristics, current status, and expectations. It is also desirable to obtain some confidence level for this classification and/or the ability to predict aggregate behavior among cohorts of similar individuals. As described in Section II, many neural network architectures are very well suited to developing this type of attribute to class mapping directly from information in the data sets. Again, the application of networks to this problem is very straightforward. Once a network architecture is chosen, the independent and dependent (enlist/not enlist) variables are supplied to the network which trains itself to best reproduce the observed enlistment behavior. One advantage of neural networks, as mentioned before, is their ability to develop nonlinear interactions among the independent variables. It is difficult to specify all of these potential interactions and impossible to know the functional form of the relationships. None of the studies mentioned above considered these types of interactions; and, in any case, the specific form of the interactions could not have been specified before estimation.

Reenlistment/Separation

Retention and reenlistment of enlisted airmen is one of the most heavily researched areas in the Air Force personnel system. These models are particularly relevant to the current research for two reasons. First, extensive data sets have already been prepared and this significantly reduces the cost of applying neural networks to the problem. Second, the breadth of research in the area has enabled researchers to view the issue from many perspectives and apply several state-of-the-art statistical techniques to the problem. This breadth of techniques provides fertile ground against which to compare the results obtained with neural networks.

Most of the models in reenlistment or retention are based on entity data. Researchers attempt to explain and quantify the factors which affect reenlistment decisions made by individual airmen. Observations on the past decisions made by airmen are analyzed in the context of the airman's characteristics and the conditions facing the airman at the decision point: military pay, Air Force policies, and civilian opportunities. Some of the models also attempt to model extension behavior as either a stepwise process or a process simultaneous with the reenlistment decision. As seen with enlistment decisions, this is an archetypical classification problem, and one to which neural networks are particularly suited.

Some Specific Models

Specialty-Specific Models. The research of Saving, Stone, Looper, and Taylor (1985) is representative of the approaches normally taken in analyzing reenlistment behavior. They studied and quantified the factors affecting first-term, second-term, and career airmen making reenlistment decisions. Based on individual-level data from the Uniform Airmen Records (UAR) and the Airman Gain/Loss (AGL) files, the researchers estimated probit equations to explain the observed reenlistment behavior (see Table 1 for a list of independent variables). One

unique aspect of the research is the detail at which separate equations were estimated; most estimations were performed at the four-digit Air Force Specialty code (AFSC) level. Saving and Stone (1982) used an early version of this reenlistment model to analyze the impact of "people programs" (base of preference, joint spouse assignment, etc.) on first- and second-term Air Force reenlistment.

Eighty-five of the original equations used by Saving et al. were evaluated by Stone, Looer, and McGarity (1990b) using new data beyond the original estimation sample. Utilizing quarterly and monthly reenlistment rate projections, the research found that the equations consistently under-predicted reenlistment rates in about one-third of the Air Force specialties (AFSCs). This led to a respecification of the models (see Table 1) and the addition of an exponentially declining time variable to reflect changing attitudes toward the military after the mid-1970s. In addition, the employment rate factor was modified to include two terms: employment rate and employment rate squared.

An approach similar to the models reviewed above was taken by Lakhani, Gilroy, and Capps (1984) to investigate reenlistment in the Army. Reenlistment decisions for individuals from 98 Military Occupational Specialties (MOSs) receiving SRBs were taken from the 1980 and 1981 Enlisted Master Files (EMFs). These 98 MOSs were then aggregated into 15 Career Management Fields (CMFs). Separate logit equations were estimated on each of the CMFs. As can be seen in Table 1, Lakhani et al. used a much smaller set of explanatory variables.

Terza and Warren (1986) extended the Lakhani et al. model to include a simultaneous estimation of the reenlistment/separation/extension decision of Army soldiers using a reduced-form trinomial probit model. In addition to testing trinomial probit, Terza and Warren also estimated multinomial logit equations for 15 CMFs. Although specification tests indicated that the trinomial probit estimator was appropriate, the researchers found that out-of-sample predictions were inferior to those produced by a simple logit model.

ACOL Models. Warner and Goldberg (1983) developed the ACOL model while analyzing the reenlistment decisions of Naval personnel. This model attempts to bring all of the pecuniary factors affecting an individual's reenlistment decision under the umbrella of a single value based primarily on the present value of potential income streams. The military income stream includes an accounting for RMC, SRB, and retirement pay, with explicit accounting for tax effects and expected promotions. A completely separate OLS equation was estimated to predict civilian earnings. In a similar vein, Black and Ilisevich (1984) developed an ACOL-based separation model using survey data covering all four DOD services. Their estimation data set was based on a 1-year DOD survey performed in 1978. Black and Ilisevich estimated a separate enlisted personnel equation for each service and an aggregate DOD enlisted personnel equation. As seen in Table 1, additional information was available on the survey instrument to provide a better accounting of individual taste for the military.

The ACOL-2 model was developed by Smith, Sylvester, and Villa (1989) to include a structural linkage between first- and second-term reenlistment behavior in the Army. They sought to measure the impact of first-term ACOL and other first-term independent variables on second-term reenlistment. Their findings indicated that the effect of first-term conditions on second-term reenlistments was dominated by actual conditions at the second-term decision point.

TABLE 1. INDEPENDENT VARIABLES USED IN FIRST-TERM
REENLISTMENT/RETENTION MODELS

Independent Variables	Saving et al. 1985	Stone et al. 1990b	Saving & Stone 1982	Lakhani et al. 1984	Terza & Warren 1986	Warner & Goldberg 1983	Black & Hiserich 1984	Smith et al. 1989	Kohler et al. 1988	Carter et al. 1987
Demographics										
Minority	X	X	X					X		
Black				X	X		X	X		
Hispanic							X	X		
White & female										X
Black & male										X
Female	X	X	X			X	X			
Single or married indicator	X	X	X			X	X			X
Age less than or equal 17 years	X									
Age greater than or equal 19 yrs.	X									
Age							X			
Two or more dependents	X	X	X							X
Number of dependents				X	X			X		
Spouse in military							X			
Spouse in civilian job							X			
Education										
Education Level							X	X		
High school or better	X	X	X							
Not a high school graduate										X
Some college education										X
Aptitude										
AFQT I or II	X	X								X
AFQT IV or better										X
AFQT score or percentile										
Mental category I to IIIA				X	X			X	X	
Pecuniary										
SRB	X	X								
Prior- & post-month SRB		X			X					X ¹
Present value of military income	X	X								
Present value of civilian income	X	X								
Ratio of RMC to civilian wages			X	X	X				X	
ACOL (or version of ACOL)						X	X	X		
Financial assets							X			
Employment or unemployment rate	X	X	X			X		X ²	X	
Unemployment rate in home state				X	X					
Employment rate squared		X								
Institutional										
Induction rate	X	X	X							
Quarterly force level			X							
Percent manning attained			X							
In Air Force people programs			X ³							
Military educational benefits							X			

TABLE 1. INDEPENDENT VARIABLES USED IN FIRST-TERM
REENLISTMENT/RETENTION MODELS
(Concluded)

Independent Variables	Saving et al. 1985	Stone et al. 1990b	Saving & Stone 1982	Lakhani et al. 1984	Terza & Warren 1986	Warner & Goldberg 1983	Black & Illerich 1984	Smith et al. 1989	Kohler 1988	Carter et al. 1987
Grade				Military Aspects				X		
Year of Service (YOS)	X	X								
Term of enlistment (TOE)								X		
DOD service (Navy, etc.)							X ⁴			
Military specialty										X ⁵
Promotion eligibility rate								X		
Gender/race/AFSC combinations										X ⁶
TOE/YOS combinations										X ⁷
Age less than 18 & 6 year TOE										X
				Other						
Quarterly Attitude (strictly a function of time)		X								
Tastes for the military Fiscal Year past 1982		X					X ⁸	X		

¹ Kohler's survivor model contained eleven separate coefficients for SRB, one for each time period of the survivor curve.

² Unlike the other models, Smith et al. used unemployment at time of enlistment; for all others, time of decision is used.

³ Dummies for base of preference, join spouse and humanitarian assignment, and an indicator for any people program.

⁴ One-digit DOD occupation codes.

⁵ Separate dummies for 5-digit AFSC if over 50 cases for the AFSC; otherwise, 2-digit career field if over 50 cases.

⁶ Three dummies: female in support and administration, female in unknown specialty, and black male in support and administration.

⁷ Five dummies: TOE 4 and YOS 2, TOE 6 and YOS 1, TOE 6 and YOS 2, TOE 6 and YOS 3, TOE 6 and YOS 4.

⁸ From four surveyed variables involving military/civilian scaled comparisons on: having a say, interesting work, job security, and job location.

Models Supporting EFMS. A set of Air Force personnel loss models similar to the individual reenlistment models already discussed has been developed in support of the Enlisted Force Management System (EFMS) by Carter et al. (1987). These equations were all estimated using simple OLS on binary dependent variables (linear probability model), and the impact of all independent variables (including SRB and RMC) was assumed to be the same across all specialties. A single estimation was run across all AFSSs. As can be seen in Table 1, the retention equation contains indicator or dummy variables for each AFS. This allows for a specialty-specific base retention rate; however, no other model parameters are allowed to vary among the specialties. Contrary to the experience of Saving et al., Lakhani, et al., Warner and Goldberg, and Kohler, the Carter et al. results showed no statistical evidence that the effect of SRB varied across specialties.

The performance of the Carter et al. models was tested by Abrahamse (1988) by embedding the models into an extensive EFMS inventory planning model and comparing the resulting projections against those of the Airman Loss Probability System (ALPS). The projections from the ALPS system are based solely on the behavior of inventory groups in the year preceding a projection. The ALPS system uses no regressions and does not consider any exogenous factors such as SRB or RMC changes. Abrahamse's comparison produced mixed results. In

general, the EFMS models performed somewhat better than ALPS, but failed to fully account for changes in the decision environment.

Other Models. Using a much different approach, Lakhani (1987) sought to measure the impact of RMC and SRB on quit rates while accounting for the simultaneous impact of quit rates on SRB. Toward this end, he estimated a pair of simultaneous equations between quit rates and SRB using three-stage least squares. Kohler (1988) also took a quite different approach to the analysis of retention. He estimated survivor functions (Kalbfleisch & Prentice, 1980) for 15 primary occupational specialties and five DOD occupation codes (Table 1 contains the list of independent variables used in the models).

Reenlistment Assessment

Researchers have tested many different specifications of reenlistment models. This variety of specifications stems primarily from a problem endemic to behavioral modeling. The independent variables a researcher would like to employ are either unobservable or difficult to quantify fully. For example, individual taste for the military style of life would be a very relevant variable; however, this is not an observable quantity. Researchers attempt to capture some of this variable's impact by including other (hopefully related) variables such as race, gender, age, and number of dependents.

An example of a variable that is extremely difficult to quantify is the present value of a military career versus civilian employment. Obviously, civilian and military wages are components of the variable, as is SRB. However, personal discount rates are likely to vary among socio-economic groups and across genders. The same can be said for employment rates, which affect the expected probability of earning a civilian wage. In practice, all of these component variables are included in most specifications to account for as much of the "desired" variable's effect as possible.

As can be seen in this simple example, some of these variables (gender and race) appear as components of both "desired" independent variables. These component variables actually represent two different desired variables, and each of the desired variables may have nonlinear effects on reenlistment. In addition, the effects of SRB and military compensation inextricably mix with the values of these component variables. For example, the coefficient on the present value of civilian earnings may not be easily interpreted if gender is included in the equation and gender influences personal discount rates. In this case, the coefficient on the gender indicator would contain both gender and unspecified present value effects. Likewise, the coefficient on civilian wage would reflect some impact of the unmodeled differences in personal discount rates. Thus, the simple interpretation of coefficients from linear (or simple nonlinear) models may be severely clouded by unmodeled interactions and multiple contributions among the included independent variables.

On a related topic, most of the reenlistment models discussed employed separate reenlistment equations estimated for each specialty or group of related specialties. In most cases, the researchers found the impact of many independent variables to be substantially different across these equations. The conceptual argument usually employed to justify estimation of separate equations involves the differing civilian labor markets facing airmen in dissimilar specialties. However, this argument can easily be extended to different races or genders. Each of these groups faces a somewhat distinct labor market and several other unique conditions. If SRB levels impact on the specialties differentially, it is also quite likely that they impact these race and gender cohorts differently. The same argument holds for individuals with differing aptitudes or educational backgrounds. The possibilities for differing impacts, conditional impacts, and interactions among the inputs are countless. It is almost inconceivable that any simple linear specification of a model in this environment accurately mirrors the underlying complexity of the

relationships. Without this accurate reflection, the relationships estimated by one of these models are suspect.

Like many other researchers, Saving et al. (1985) originally found that their estimates of the impact of changes in the employment rate were unstable and prone to become positively related to reenlistment. This was in contrast to the a priori theoretical expectation that higher employment rates should increase an airman's expected civilian earnings and drive down reenlistment rates. Stone et al. (1990b) found that the additional flexibility obtained from adding the squared term kept the impact of employment within its theoretically expected range (negative). If the combination of a linear and squared term represents the "true" relationship between reenlistment and the employment rate, the functional form of the original reenlistment equations were misspecified. This misspecification (an unmodeled nonlinear relationship) would cause all coefficients from the model to be biased—particularly the coefficient on employment rate.

Neural Network Systems and Reenlistment Models

The employment specification problem encountered by Saving et al. demonstrates a domain in which neural networks are particularly appropriate. Several network architectures are capable of "discovering" such a nonlinear underlying relationship directly from the data set. The researcher is not required to search through a potentially enormous set of functional forms. This is especially beneficial in that the search process itself may destroy the validity of the statistics produced for the final model (see Leamer, 1978). Neural networks do have some disadvantages for this type of modeling, however. They do not produce coefficients that are directly interpretable. Because the network can produce complex and intermingled relationships, the behavior of the network must be examined over relevant ranges of inputs to determine the effects on reenlistment. However, if a linear model is misspecified, there is little use in attempting to interpret its biased coefficients. A second disadvantage to the network approach regards statistical testing of the model. Neural networks utilize many weights to capture the relationships in a model. There is no neural network analogue to the coefficient standard errors usually provided by regression techniques. Although it is possible to compute some statistics of this sort using resampling methods, most neural network models are validated against separate holdout samples.

The simultaneous reenlist/separate/extend decision examined by Terza and Warren (1986) provides another example for applying neural networks. All three of the neural networks described in Section II include these multi-class decisions in their general architectures. In all three cases, the only visible change to the architecture is the addition of an extra output neuron. As mentioned in Section II, back propagation can perform vector-to-vector mappings. In this case, the input vector is merely the set of independent variables and the output vector becomes three neurons representing the three possible decisions. The PNN architecture simply estimates three underlying PDFs rather than two. Similarly, each LVQ reference vector can be labeled with one of three decision paths. In all cases, the simultaneous effects of all inputs on all potential decisions are considered.

Although neural network techniques are directly applicable to models with the ACOL construct as an input, ACOL runs contrary to the strengths of neural networks. Information that might be constructively used in developing nonlinear relationships has already been embedded and lost in a linear aggregate. If the ACOL construct has been properly constructed, a neural network will be able to "learn" the linear or nonlinear relationship between ACOL and reenlistment. However, if this relationship is linear, a standard estimator will perform as well as the neural network. If the relationship is nonlinear, is the linear ACOL construct likely to accurately represent the "true" pecuniary horizon facing the decision maker? Assuming the ACOL construct could use some adjustment from other demographic, aptitude, and education variables, what

interpretation can be placed on the impact of ACOL alone? Neural networks can be expected to perform better if provided with all of the information so that any required nonlinear relationships can be developed. If ACOL is included as a neural network input, many of its components, as well as factors found important in other research, should also be included. In this manner, the network can adjust for any biases built into the ACOL construct.

Neural networks can also be applied to model-seeking problems. Carter et al. (1987) found at least 10 significant interaction terms which were included in their first-term continuation equation (see Table 1). Although it affects the statistical interpretation of the final coefficient standard errors, this type of model-seeking or specification search can prove fruitful in developing realistic models. In general, the exact form of any relationships and interactions cannot be specified before estimation. The widespread use of linear or simple nonlinear functional forms results more from computational simplicity than theoretical imperative. In addition, the parameters of the simple linear specifications are easy to interpret. As seen in Section II, neural networks offer a solution to this model-seeking problem. Because they inherently allow for the formation of nonlinear and interacting relationships, neural networks provide a method of seeking the model form supported by the empirical evidence in the data set.

Prior Service

Prior-service accessions comprise a much smaller component of the force than non-prior-service accessions, and they have traditionally had less impact on force size and management than on reenlistment rates. There have been correspondingly fewer studies of this manpower market. Stone and Saving (1983) undertook one of the few studies of this area. These researchers modeled the Air Force prior-service market by Break-in-Service (BIS) groups. For each of five BIS groups, they estimated a separate equation (OLS and two-stage least squares) containing independent variables for unemployment, RMC-to-civilian-wage ratio, recruiting effort, time of year, and the distance to prior-service recruiting goals.

Inventory Planning Models (IPMs)

As mentioned earlier, inventory models typically serve one of two purposes: long- to middle-range planning or short-range programming. In general, IPMs attempt to model and project some portion of the personnel and information flows shown in Figure 7. These models typically treat the personnel inventory as either a matrix of relevant personnel cohorts or as a collection of separate individuals (entities). Most IPMs use some form of estimated reenlistment or retention equation to help project retention, and they may also include empirical models of the accession market. The results of other analyses in areas such as retraining, prior service, attrition, and extension may also be incorporated into an IPM. These empirical or analytic results are usually combined with a base personnel inventory (known "personnel system constraints") and policy factors to develop a system of personnel stocks and flows. Virtually all IPMs exist as computer-based, discrete simulations utilizing varying amounts of analytic results on components of the personnel system.

Cohort-Based Inventory Models

An example of an IPM is the Air Force Retention Analysis Package (AFRAP), which serves primarily to analyze the impact of various factors on retention (Stone, Wortman, & Looer, 1989). This package is basically a computerized implementation of the reenlistment results of Stone et al. (1990b). All of the occupation-specific equations from Stone et al. are combined with a model of retention to produce a small IPM. The impact of changes in pecuniary factors and AFS composition (demographic attributes, education level, etc.) can be evaluated on both

short- and long-term retention. The package also has the ability to solve for SRB levels required to obtain a specified retention rate given the economic conditions and AFS composition. AFRAP does not attempt to model accessions, retraining, or assignments.

A second example of an IPM spectrum is the Enlisted Force Management System (EFMS), which seeks to model virtually all aspects of the Air Force personnel system. As originally described (Carter, Chaiken, Murray, & Walker, 1983), EFMS sought to support most force management activities: requirements determination, personnel planning, authorizations management, and programming. To support all of this analysis, EFMS was to include three mutually consistent IPMs: short-term to address the remainder of a year, middle-term for monthly projections up to 7 years, and long-term for monthly projections for an arbitrary number of years (Carter et al. 1987). Though EFMS is not yet completed, several components will be discussed below.

The EFMS Bonus Effect Module (BEM) is based on the EFMS middle-term loss equations (Carter et al., 1987) and allows analysis of bonus effects without running the large, entity-based EFMS IPM (Carter, Skoller, Perrin, & Sakai, 1988). Similar to AFRAP, BEM is designed to perform analysis on a single selected AFSC and produce inventory counts by YOS. BEM provides more cost information than AFRAP but assumes that economic conditions are stable. The primary policy level available to the BEM user is bonus.

Michelson and Rydell (1989) produced another IPM based on the EFMS middle-term loss equations of Carter et al. (1987). The Aggregate Dynamic Model (ADAM) is a cohort- or cell-based IPM which projects the enlisted force personnel inventory along three dimensions: grade, YOS, and TOE. In addition, the model retains another inventory dimension, years to end of term (YETS), to provide accounting required by the loss equations. ADAM requires inputs on economic conditions, accessions, some separations, and promotions. Unlike AFRAP and BEM, ADAM does not provide AFS-specific projections. It does, however, produce aggregate inventory projections in a more accessible format and provides for a more complete accounting of inventory flows.

The Enlisted Policy Planning System (EPPS) represents another inventory model based largely on empirical reenlistment or loss equations (Syllogistics, Inc. & RRC, Inc., 1989). The system was designed as a planning model for policy analysis to determine the effects of program and policy changes. EPPS adds a 4-digit AFSC to the breakdown along the YOS, grade, and TOE dimensions found in ADAM. The primary behavioral models consist of the reenlistment/separation equations estimated by Stone et al. (1990b). Inputs into the EPPS model include economic conditions, AFSC/grade manning authorizations, and personnel policy variables.

Other Inventory Models

The Airman Loss Probability System (ALPS) produces loss probabilities for each airman in the final UAR. These probabilities are normally used to derive loss rates and reenlistment rates for each AFSC/grade/YOS cohort. Unlike the behavioral reenlistment models, ALPS bases these rates solely on the two most recent UARs and a transaction file containing promotions, demotions, gains, and losses. The resulting transition rates are based on the observed behavior in the cohorts and do not explicitly account for economic or policy factors such as unemployment, SRB, RMC, etc. Despite its simplicity, ALPS rates have been utilized in several IPMs: the Airman Inventory Projection System (AIPS), the Airman Force Program and Longevity Model (AFPAL), and the Dynamic Model. For the purposes of inventory modeling, the reenlistment efforts of Saving, Stone, Lakhani, Looper, Goldberg, Black, and others serve to improve the foundation on which many inventory transitions are based by adding new information. Neural

networks may further improve this base by allowing unique and meaningful combinations of this information.

Fernandez, Gotz, and Bell (1985) developed a model of airman retention based on the dynamic model of Gotz and McCall (1980, 1985). Though not a complete IPM, the dynamic retention model explicitly incorporates the sequential decision-making process involved in making multiple reenlist/separate decisions. It also takes account of tastes and past conditions by explicitly modeling the entire sequential decision process.

Stone, Saving, Turner, and Looper (1990a) developed a set of four equations which, though not a true inventory model, describe aggregate Air Force accessions (prior- and non-prior-service) and reenlistments (first- and second-term). In addition to estimating the equations by OLS, Stone et al. estimated the entire system using a generalized least squares (GLS) estimator. A specification test between the OLS and GLS models indicated a correlation among the error terms and a significant difference in the coefficients across the two models. The GLS estimator performed better in a simulation of the two models over a time period prior to the estimation sample. Conversely, the OLS estimator performed better in a post-estimation time period.

Other inventory models include the Integrated Simulation Evaluation Model Prototype (ISEM-P) model developed by Rueter, Kosy, Caicco, Laidlaw, and Looper (1981). This model was designed to predict personnel system implications of changes in policy information control (PIC) policies and procedures and the impact of changes in national labor markets. The Career Area Rotation Model (CAROM) represents a very different approach to inventory modeling (Looper, 1979). The goal in CAROM is to optimize enlisted assignments on a monthly basis using an entity-based model for a single AFSC. The model uses Monte Carlo techniques and linear programming to allow policy gaming for planning purposes.

Neural Networks and IPMs

Neural networks could be easily incorporated into a system such as AFRAP. The networks would simply replace the probit estimations currently used to model each AFS's reenlistment decision. The potential benefits are the same as those presented in our earlier discussion of reenlistment models. The network models allow nonlinear impacts and interactions among the input factors. In essence, the neural networks could capture more complicated and potentially more realistic models of the process. As with AFRAP, the primary application for neural networks in BEM and ADAM would involve the development of more complex loss functions. The primary use of neural networks in EPPS would be to improve the behavioral equations and perhaps analyze some of the assumed fixed flow rates.

Without extensive theoretical groundwork, neural networks could not be directly applied to the dynamic retention model. The dynamic retention model's estimation and simulation methods are specifically tailored to its sequential structure and the specific derivation of its aggregate present value measure. However, neural networks can capture both the sequential nature of the decision-making process and the generation of a meaningful composite variable. The sequential decision process is addressed using a recurrent form of back propagation (Elman, 1989, 1990). Meaningful composite variables are derived by filtering several input variables through a single neuron. This neuron will then represent the "best" nonlinear combination of the chosen inputs for predicting the observed behavior (separation/reenlistment). "Best," in this context, means simply that composite variable which can be used to produce the closest sum of squared error (or maximum likelihood) fit to the observed airman behaviors. In this manner, and unlike ACOL, the composite variable produced by the network is not restricted to a prespecified functional form.

Neural networks, and particularly back propagation networks, are directly applicable to the simultaneous accessions/retention model of Stone et al. (1990a). In this case, the network outputs are merely the two accession rates and the two reenlistment rates from the original model. Although it is possible to develop separate networks for each equation, this model would probably be best treated with a single network having four output neurons. All of the independent variables used in the estimation would serve as inputs, and the network would develop an internal model of the system in its hidden-layer neurons. As with the other models, the ability of the network to generate nonlinear relationships could be of considerable importance to the simultaneous accession/retention model. A potential addition to the model involves the use of Elman's simple recurrent network (SRN). With this network, the representation developed in the hidden layer is used as network input for the ensuing time period. In this manner, the network is able to develop temporal relationships and account for sequential adjustments in the system.

Other Personnel-Related Models

The models reviewed above are drawn primarily from areas applicable to military personnel inventories, and they focus on the primary personnel flows shown in Figure 7. Personnel decisions must be made in many other ancillary areas, and special programs must be administered. The policies adopted in these areas and programs can often benefit from the application of analysis and modeling tools.

Armed Forces Health Professions Scholarship Program (AFHPSP)

One such area involves the AFHPSP. McGarrity (1988) developed a policy-specifying model (see Fast & Looper, 1988) that could be used to assist a review board in selecting candidates to this program. The inputs to the model consisted of 13 factors such as academic potential, military experience, and personal experience. These factors were utilized to develop a standard hierarchical policy-specifying model based on the input of subject-matter experts (SMEs). The SMEs supplied pairwise relationships between the factors, and payoff values for the resulting combinations.

Neural Networks and the AFHPSP

Applying neural networks to this problem would produce results similar to policy capturing (Fast & Looper, 1988). A network could be trained using the 13 inputs for each applicant and the review board's score for the applicant. The resulting model would be analogous to a nonlinear policy-capturing model, which seeks its own nonlinear specification. Factors or combinations most important to the review board in rating an applicant could be located by analyzing the resulting network. These combinations of factors would be determined by the board's observed actions rather than by surveying their opinions. In addition, it is possible to apply the policy-capturing technique of interrater clustering to the hidden nodes in a neural network. In this manner, if separate networks are estimated for each board member, it becomes possible to identify rating patterns which differ among board members.

Recruiter Assignments

The recruiter assignment model developed by Looper and Beswick (1980) might be considered an accession model. However, its primary goal is to determine the optimal allocation and assignment of recruiters. The Looper and Beswick model uses a nonlinear estimation equation and dynamic programming to maximize the number of recruits subject to a fixed number of

recruiters. As with many of the other models discussed, the primary use of neural networks in this application would involve the development of a more flexible nonlinear function.

IMPLEMENTING NEURAL NETWORK PERSONNEL MODELS

As discussed in the previous sections, neural networks have several potential applications in personnel modeling. This potential should be evaluated in at least two different areas: reenlistment analysis and inventory projection. These areas represent some of the more important personnel issues and also very different challenges as empirical problems. Reenlistment analysis is representative of many classification problems in the Air Force. It remains one of the most thoroughly analyzed personnel issues. Alternately, inventory projection involves analyzing and forecasting personnel inventories.

Most current neural network applications are directed toward relatively small, well-understood problems (see Wiggins, 1990a). These types of problems have been chosen for two primary reasons. First, neural networks can be computationally intensive and require long simulation times on standard serial computers. Most large networks implemented on serial hardware require exponentially longer training times than do small networks. Though hardware solutions are becoming available to address this problem, they are currently rather costly. Second, the performance of any model on large problems is much more difficult to assess. Most research projects have been aimed toward testing neural network capabilities in various problem domains. If the network's performance relative to other methods cannot be established, its capability is difficult to assess. Model assessment is critical to most neural network research. Although theoretical results have placed high upper bounds on the capabilities of neural networks, these results have yet to be extended to training and training dynamics. Despite a host of promising empirical results, the uncertainties about training make validation and assessment of neural network models very important.

For these same reasons, preliminary personnel research using neural networks should be kept to a reasonable scale. In the two tasks addressed below, an attempt has been made to balance attention to substantial problems with considerations of meaningful assessment and cost of performance. Each task addresses important personnel areas, while retaining a modest scope. More traditional personnel models are available against which the performance of the neural network models could be compared. If these preliminary network models exhibit superior capabilities, larger models requiring hardware support might be attempted. However, many other moderately sized personnel applications could benefit from smaller software-based neural networks.

A final consideration in selecting problems to be modeled involves data availability. As discussed in Section II, most neural network architectures require more information than traditional techniques require to produce a model. With most statistical techniques, the functional form of the model is imposed by the researcher. Because neural networks infer the form of the model from relations in the training data, sufficient data must be available to make meaningful inferences about the underlying process structure. More information is required from the training data because the researcher does not supply prior information in the form of an imposed model structure.

Reenlistment Model

The model-seeking capabilities of neural networks make them particularly suited to individual reenlistment modeling. As discussed in Section III, rarely can the functional form of a behavioral reenlistment model be specified by theory alone. From observed behaviors, neural networks

have the ability to directly develop internal representations of a model's form. Reenlistment models meet all of the criteria for good test model candidates:

- Reenlistment models are valuable tools in many aspects of personnel analysis and management.
- They have been extensively researched, and many state-of-the-art models are available for comparing results.
- The models are relatively small and have few enough inputs to allow analysis and evaluation of the results.
- Data on observed reenlistment behaviors are plentiful and readily available.

Model Structure

A major goal of the present research was to explore and assess alternate neural network architectures. As seen in Section II, and in Wiggins (1990a), many architectures are available and most have several variants that emphasize the solution of particular problems. In addition, modified techniques and improvements are being developed at a rapid rate. The research in reenlistment modeling should remain sufficiently flexible to allow investigation of new and promising neural network techniques. In light of this, the research should be restricted to a small set of AFSCs. These should be chosen such that the following AFSC characteristics are included: a small AFSC, a large AFSC, an AFSC receiving little or no SRB multiples over the period analyzed, and a Cronically Critical Shortage (CCS) AFSC with substantial changes in SRB. This will allow for some comparison of network models developed from large and small data sets in the same problem domain. Because the first-term equations contain the richest data and structure, only first-term reenlistments need be considered.

The neural network should be trained on continuous values underlying some of the indicator variables used in prior reenlistment studies. Use of the continuous variables removes the judgment and experience of the researcher from the specification. Because the network can develop nonlinear response surfaces, it is not necessary to impose a specific discontinuous indicator variable.

In addition to modeling the reenlistment/separation decision of eligible airmen, the neural network architectures should also be applied to extension behavior. This model more completely represents the choices facing an airman near his ETS. In this case, the decision becomes reenlist/separate/extend and the inputs remain those from Table 1. As mentioned earlier, most neural network architectures extend quite naturally to multi-class decision problems.

Modeling Techniques

Many neural network architectures are applicable to classification problems such as reenlistment decisions. All architectures discussed in Section II (back propagation, LVQ, and PNN) are particularly suited to classification and should be applied to reenlistment modeling. The strengths and weaknesses of the network architectures in this arena can then be compared against each other and against the results of probit analysis. Each of these architectures can also be used to analyze the more complete reenlist/separate/extend problem.

In all cases, any modifications or additions to the architectures which improve generalization performance should be tested. For LVQ, this will involve testing differing numbers of reference

vector neurons. For PNNs, this usually involves only the setting of the Gaussian smoothing parameter. In addition, weighting of the inputs should be employed to increase the use of information in the sample. These weights can be developed using maximum likelihood techniques and the hold-one-out sampling process described earlier. In the case of back propagation, several features designed to improve generalization should be evaluated:

- Holdout sample to stop the training process.
- Exponentially declining network weights to reduce sensitivity to noise.
- Alternative transfer functions, such as the one based on Tukey's distribution.

Data Requirements

Because the three neural network architectures require precisely the same data that Stone et al. (1990b) employed to estimate probit reenlistment models, the Stone et al. results could serve as an excellent testbed for neural networks applied to Air Force personnel system modeling. As mentioned in Section III, these data were compiled by matching the UAR and AGL files from 1974 to 1980. Each AFSC was segregated into a separate data set for estimation. Additional information was appended to the files from BLS and Census sources (civilian wages and employment rates). These data sets could be used directly to train and validate the neural network models. Although not used by Stone et al., information on extensions is also available from the AGL files.

Validation and Testing

Many validation methods are applicable to neural network and probit reenlistment models. Two distinct methods should be considered here. In the first method, a set of observations on individual airmen are randomly withheld from the training (or estimation) sample. This holdout sample would then be used to test the model which results from training (estimating) on the training sample. Predictions of the behavior of each individual in the holdout sample are made by each model, and these predictions are compared against the actual decisions observed. Each of the models—probit and neural network—produces continuous predictions which can be viewed as the probability of reenlisting. By use of a cutoff value, these probabilities may be interpreted as either a reenlist or a separate decision. For example, a predicted reenlistment probability of 0.6 is usually construed to denote a reenlistment, whereas a probability of 0.3 implies a separation.

With such binary outcomes, a simple measure of success is the hit-rate or percent of successful predictions.¹⁴ This measure was used extensively in the neural network classification literature reviewed in Wiggins (1990a). It provides an intuitive method of comparing the performance of different models against observed behaviors. The receiver operating characteristic (ROC) from signal detection theory provides another validation measure for binary outcomes (Spoehr & Lehmkuhle, 1982). The ROC is also based on prediction hits. Unlike the hit-rate, the ROC can be tuned by varying the cutoff value. Though the ROC measure has some weaknesses in this context, both of these measures should be applied to the neural network models developed. The tests will require retraining each network on the randomly selected

¹⁴All of the validation measures for evaluating the reenlistment models and the simultaneous accession/retention models are discussed in greater detail in Stone et al. (1990b).

training samples before comparisons could be made using the holdout or validation sample. Probit models should also be estimated on the training sample, with the ROC and hit-rate measures computed over the holdout sample. The probit model can then serve as the basis for evaluating the relative out-of-sample performance of the network models. These tests should be repeated for each of the four selected AFSSs.

In addition to the validation tests on a single holdout sample, the hold-one-out validation sampling described in Wiggins (1990a) could be applied to the probit and PNN models. Using hold-one-out sampling for validation allows more of the data from the original sample to be used in estimating each model. This may be particularly important for the PNN, which is estimating a high-dimensional PDF. The other neural network architectures could also make good use of any additional training observations in forming a model. However, the longer training times required for LVQ and back propagation make hold-one-out sampling unworkable for these architectures.

As mentioned earlier, each of the models produces continuous reenlistment probabilities. Because of this, their performance could also be analyzed using any of the RMSE-based measures described in Stone et al. (1990b): Thiel's inequality coefficient, Janus quotient, predicted/actuals correlation, normalized prediction error, and simulation R-squared. However, given the binary nature of the actual outcome (reenlist/separate), interpretation of these measures can be vague. Most are scaled such that a value of 0 or 1.0 implies some form of perfect prediction or complete failure to predict. However, the binary nature of the actual outcomes usually prevents any continuous output from approaching perfect prediction. For this reason, although these RMSE-based measures can actually contain more information than do the binary measures, hit-rates typically are used to evaluate binary outputs.

The second validation method is related to the use of reenlistment models in IPMs and was utilized by Stone et al. (1990b) to validate their original reenlistment equations. This method involves projecting the reenlistment behavior of temporal cohorts of decision makers. The probit equations were estimated over the 1974 to March 1982 time period. These equations were then used to project the reenlistment rates over the April 1982 to April 1986 time period. The ability of a model to accurately project the behavior of temporal cohorts is critical to its behavior in an IPM where these rates are its sole output. The neural network models could be evaluated using the same temporal sub-samples employed by Stone et al. (1990b). The temporal cohorts would be sampled quarterly over the out-of-sample time period, with the projected reenlistment rates for each quarter compared against the actual rates for the quarter. The Janus quotient, Thiel's coefficient, and simulation R-squared would be used to compare the performance of the models. In addition to these RMSE-based measures, the normalized prediction error (also RMSE-based) and the correlation between actual and predicted rates should be computed for each model.

Evaluation and Interpretation of Models

The complexity of neural network models makes them more difficult to interpret than standard parametric models. Even if the model performs well in-sample and out-of-sample, the reason for its performance and its behavior over different input ranges cannot be evaluated directly. The very aspect of neural networks that gives them a powerful analytic capability makes them rather difficult to interpret. The nonlinear and interacting relationships captured by a network are embedded within the network's weights, forming complicated composites of the inputs. Evaluating the behavior of such a network requires considerably more effort than checking the sign of a regression coefficient. However, the results of such an effort could reveal interesting structures in the underlying model. For example, Stone et al. (1990b) found that the employment rate had a more theoretically appealing impact on reenlistment if it entered the probit equations in both linear and squared forms. Carter et al. (1987) found several combinations of indicator

variables which had independent impacts on reenlistment likelihood. Certainly the neural network reenlistment models should be examined to see if these same structures emerge. Furthermore, the entire surface of each network's response surface should be searched for nonlinear impacts. The surface should also be searched for interaction areas where the impact of one variable on reenlistment is affected by the level of another variable.

In general, these types of interactions and nonlinear relationships can be found only by searching over the model's response surface. The marginal effect of changing one input while all other variables are held constant could be evaluated at any point corresponding to a set of fixed input values. This effect is simply the derivative of the probability of reenlistment with respect to a change in one input variable while all other variables are at a pre-specified point. This derivative can be derived analytically for back propagation networks by simply propagating the error all the way back to the input layer. PNN and LVQ networks require the use of numerical methods to compute the derivative or marginal effect. Still, in all three cases, the computations are straightforward.

With any of the three neural networks, these marginal effects can change from one point on the model's surface to another. For example, changing RMC by \$100 per month when unemployment is relatively low, say 6%, may have a large effect on reenlistment. Making this same \$100 change when unemployment is 20% may have very little effect. With civilian job opportunities severely limited, airmen may not require an added incentive to remain in the force. With linear models, the marginal effects are constant at all points on the model's surface. Similarly, with log-log models, the marginal percentage effects are the same at all points on the surface. The only way to introduce nonlinearities is to specify them directly in the function as did Stone et al. Likewise, the only way to introduce co-dependent or interacting effects, such as the one between RMC and employment, is to explicitly specify the form of the relationship. Only Carter et al. (1987) examined co-dependent effects, and they looked only for effects between indicator or dummy variables.

A trained neural network model does not "announce" the form and location of interactions and nonlinearities; however, the response surface of the model could be searched for such interesting features. One way to search for such features involves evaluating the marginal effect of each variable at all points on a multidimensional lattice spanning all inputs. The extent of the lattice in each input dimension could be determined from the observed range of the input or by a prior knowledge of the relevant and interesting range. This range is then subdivided into a small number of segments (usually evenly spaced), and the process is performed for each input variable. The set of all possible combinations formed by the endpoints of these segments produces a lattice in the input space. The marginal effect of each input variable is then evaluated at all lattice intersections. Although this method effectively covers the input space, it is most effective in low-dimensional spaces (i.e., when there are few inputs). In high-dimensional spaces, the lattice method suffers from exponential increases in the number of points which must be evaluated. For example, with 25 inputs and only 3 lattice points in each dimension, over 840 billion points must be evaluated.

An alternative to the lattice method in high-dimensional input spaces involves evaluating the marginal effects at each point in the training and/or validation sample. In this manner, the density of sampling for the search is determined directly by the density of the input data. With the lattice method, many of the spaces searched may contain few, if any, individuals. By using the sample points, the search is directed toward areas where large numbers of decision makers tend to cluster. As a secondary effect, the search focuses on those areas where the model could be expected to perform best. Almost any estimation method, including neural networks, produces its most generalizable predictions in those areas of the input space with the highest exemplar density. Because of the fairly high-dimensional nature (18 inputs) of the Stone et al. model, this method of searching for interesting features is expected to be quite useful.

Given the computing power required to evaluate a network model in this manner, only the models which perform best with respect to the validation criterion would be evaluated. Those evaluated should include at least one model from each network paradigm. All nonlinear and co-dependent marginal effects from each model should be reported and compared. Specifically, the relationship between employment rates and reenlistment should be evaluated and compared to the nonlinear relationship found by Stone et al.

Inventory Model

The second model to be addressed using neural network techniques is a projection of inventory flows, which could be extended into an aggregate IPM. The model is small enough to support an IPM whose results would be easier to evaluate than those of a disaggregate IPM. Most IPMs require extensive analysis to provide any information on their performance. Even then, their disaggregate nature could make the interpretation of results difficult (see Abrahamse, 1988). In addition, the complexity of simulating with most IPMs, and the data required to perform a projection, typically limits validation tests to one or two periods. This is scant information upon which to base validation conclusions. It is hoped that an aggregate IPM will prove more tractable; however, as discussed below, even this simple IPM poses some problems of scale. In general, the other criteria for model selection have been met: Preliminary results from the model can be compared with those from another model, and a reasonably large training sample is available.

An excellent candidate IPM is the aggregate accession/retention model (AARM) of Stone et al. (1990a). To utilize this IPM, a neural network model could be developed which directly parallels the AARM. This network model could then be extended to account for more inventory flows and YOS cohorts. Finally, the resulting network model could be built into an IPM which projects aggregate force levels.

As shown in Table 2, the AARM is composed of four equations: NPS accessions, PS accessions, first-term reenlistment, and second-term reenlistment. The model was estimated using GLS on monthly data from October 1979 to September 1987. These same data could be used to develop the neural network models and IPM. As with the AARM, the January 1979 to September 1979 data and the October 1987 to September 1988 data should be used to validate the resulting models.

Initial Network Model

The initial neural network model should use exactly the same inputs and outputs as those used in the original AARM. As seen in Table 2, the input variables include measures of recruit quality, wait-time in the DEP, civilian employment, relative military/civilian wages, early outs, eligible decision makers, force-level goal, and accession goals. The network model would be trained by back propagation on a network using the 15 inputs used by AARM and having four output neurons (each representing one of the four AARM dependent variables). Techniques for improving the generalization of back propagation network models should also be applied to this problem.

Once the network model has been trained, the predicted accessions and reenlistment rates should be compared against the actual rates (both in-sample and out-of-sample). Again, all of the continuous validation measures mentioned previously should be applied to the comparison. These measures could then be compared to the same measures computed for the original AARM. As with the reenlistment network model, this network flow model should be evaluated to search for interactions and nonlinear relationships. These relationships would be particularly interesting if the network model displays superior out-of-sample performance. In addition, the

range of marginal effects of the independent variables on each of the outputs should be computed over all training observations. The distribution of these effects could then be compared with the static GLS regression coefficients.

TABLE 2. SIMULTANEOUS ACCESSION/RETENTION EQUATION SYSTEM

Right-hand-side variables	Structural Equation			
	Accession Rate	Prior-Service Accession Rate	First-term Reenlistment Rate	Second-Term Reenlistment Rate
Ratio of AFQT Categories 1 or 2 to all other accessions	X			
Average time in Delayed Enlistment Program (DEP)	X			
Civilian employment rate	X	X	X	X
Ratio of military to civilian wages	X	X	X	X
Number of Air Force recruiters	X	X		
Force-level goal	X			
Accession goal	X			
Prior-service accession goal		X		
Ratio of eligible to ineligible decision makers (first-term)			X	
Ratio of eligible to ineligible decision makers (second-term)				X
Number of first-term early outs			X	
Number of second-term early outs				X
Quarterly indicators	X	X	X	X

The simple recurrent network (SRN; Elman, 1989) provides another interesting method of modeling the AARM outputs. As discussed earlier, this modification of back propagation could incorporate sequential effects into its structure. It is quite likely that temporal adjustments are being made in the enlisted inventory at the monthly level. If so, and these adjustments have a regular structure, the SRN may be able to capture some of the system's dynamics. The resulting model should be validated and compared against the results of the standard back propagation model and the original AARM.

As originally specified, the AARM does not project sufficient information for an aggregate IPM. This structure should be extended to provide for projections of attrition and retirement. The same inputs could be used, but the network will now have six outputs: two reenlistment, two accession, one attrition, and one retirement. If possible, a YOS distribution should also be tested as input to the model. This distribution would provide some information on retirement eligibles and the number of airmen in high-attrition YOS.

If the network models are successful in projecting aggregate inventory flow rates, they could be extended to output a complete set of flows required to project a reasonable aggregate inventory model. The structure of the inventory would be kept as simple as possible, yet retain information necessary to track the aggregate inventory as it ages and cohorts approach

decision points. The inventory could be dimensioned along YOS, YETS,¹⁵ and in-extension¹⁶ cohorts. Using 31 YOS, 6 YETS, and 2 in-extension inventory dimensions yields an inventory representation with 372 cells. The number of accessions by month could be tracked to allow for monthly aging of the inventory.

The network model could use the same 15 aggregate inputs from the AARM; however, loss and extension rates would be output for each appropriate inventory cell. Reenlistment rates could be projected for 27 YOS and both in-extension cohorts, for a total of 54 rates. Attrition rates could be projected for each inventory cell (372 rates). The last 11 YOS would be considered retirement eligible for 6 YETS and 2 in-extension categories (132 rates). Finally, extensions would be allowed in 124 of the inventory cells. In all, the network model would project 682 inventory flow rates on a monthly basis. In addition, the model would continue to project aggregate PS and NPS accessions.

Little change would be required in the structure of the neural network to accommodate this expanded model. In place of 4 output neurons, the network would have 684 output neurons. Despite the simplicity of the model, this network would become quite large. It would be considerably larger than any of the networks considered in the applications reviewed in Wiggins (1990a). This model would provide a test for the ability to scale network solutions to problem domains with many simultaneous outputs and relationships. The scale of the model is at the limit which can be reasonably addressed with software simulators. Any larger model would likely require hardware support during its training phase.

Data for all of the flow rates could be derived from the AGL. The IPM could be treated as a standard discrete monthly model where the neural network controls all of the flow rates. Aging would be the only inventory flow not controlled by the network model. In addition to the aggregate inputs from AARM, representations of the existing inventory should be considered as inputs to the model.

Despite the YOS inventory breakdown, the model described above is still primarily an aggregate IPM. No cell-specific information is provided upon which to base the projected loss rates for individual cells. The use of cell-specific information should be explored. In particular, YOS-specific average RMC and SRB values could be derived from UAR counts and military pay tables. The airmen inventories in neighboring cells are another potential source of input. This cell-specific information would be provided to each output neuron through an extension of the back propagation architecture. Each set of outputs for a given inventory cell would contain a sub-network which processes only cell-specific information. This sub-network could be combined by the output cell with information from the aggregate input network. In this manner, each inventory cell has both aggregate and local factors which influence the flow rates affecting the cell.

A more complete model would require some measures of Air Force demand for personnel in each cell, such as authorization or manning requirements. Authorization and manning information would be difficult to collect for the long time series required. This process should be undertaken if the results from this task are extended to a larger inventory model.

¹⁵ As mentioned earlier, YETS is used here to represent years to end of term of service. It measures the time remaining before a reenlistment decision must be made.

¹⁶ In-extension merely designates whether the airman is currently in an extension to a prior term of service or in a "new" term of service. It can assume only two values.

Validation

The validation of the resulting IPM should be based primarily on its ability to project aggregate inventory stocks and flows. Because the inventory could be used recursively in a projection, the IPM can perform multi-step inventory projections. The model would require only a knowledge of the 15 independent variables and military pay tables for each projection period. This would allow two techniques to be applied when validating the IPM: one-step projections, and multi-step projections.

One-step projections are made such that the inventory on which a projection is based is the actual inventory before the 1-month projection. Using the Stone et al. (1990a) data, a one-step projection could be made for each month of the in-sample and out-of-sample data sets. The resulting 96 in-sample and 21 out-of-sample projections should be evaluated using RMSE-based validation techniques. Separate and joint validation measures should be computed for the in-sample and each out-of-sample (pre- and post-estimation) time period. Validation measures should be computed for each of the aggregate flows: PS accessions, NPS accessions, reenlistments, extensions, attritions, and retirements. In addition, validation statistics could be computed for the total inventory level. It would also be possible to compute validation statistics for individual inventory cells and their associated flow rates.

Multi-step projections require that the model continually operate from the same inventory. An actual inventory is provided at the start of the projection, but each successive projection is based on the inventory forecast from the last time period. This type of projection allows each forecast error to become built into the next period's forecast. Primary concerns addressed by this type of validation are model stability and sensitivity to errors or starting conditions. By starting the model from each sample period and projecting over several years, both the stability and sensitivity measurements could be addressed. The projections for a single period in time could be evaluated when the projection begins at differing starting points. By observing the model's behavior over long multi-step projections, its stability could be assessed.

The inventory model should be evaluated using all of the validation measures. Relative performance on aggregate reenlistment and accession rates between the network and AARM models should be compared. The accuracy of the final network IPM in projecting stocks and flows should be appraised using both one-step and multi-step projections. Stability and the ability to adjust to initial conditions should be assessed with multi step projections. All of this information should be evaluated in conjunction with the computation requirements of the neural network model. If the model's performance is acceptable, prospects for expanding the neural network IPM to a disaggregate inventory could be assessed.

CONCLUSIONS

Neural networks exhibit several theoretical and practical capabilities that are very attractive from a data analysis and model-building perspective. Primary among these capabilities is the ability to detect arbitrarily complex, interacting, and nonlinear relationships among the factors of a particular model. In addition, a review of the neural network literature (Wiggins, 1990a) reveals that neural networks have demonstrated substantial success in areas currently dominated by traditional statistical techniques.

Many areas of personnel analysis and management may benefit from the richer and more complex models offered by neural network methods. To assess the potential for applying these promising new techniques to personnel research, several test models should be developed in areas having existing models based on more traditional techniques. Comparisons between the behavior of the existing models and their neural network counterparts will provide some objective measures of the performance of neural networks for personnel and manpower analysis. The

extensive amount of data available in most areas of the personnel field offers many possibilities for developing rich and complex models directly from the information available in observed behaviors.

REFERENCES

- Abrahamse, A.F. (1988). *Middle-term disaggregate loss model test and evaluation: Description and results*. Santa Monica, CA: The RAND Corporation.
- Anderson, J.A., & Rosenfeld, E. (Eds.). (1988). *Neurocomputing foundations of research*. Cambridge, MA: The MIT Press.
- Ash, C., Udis, B., & McNow, R.F. (1983). Enlistments in the all-volunteer force: A military personnel supply model and its forecasts. *American Economic Review*, 73(1), 145-155.
- Ash, T. (1989). *Dynamic node creation in backpropagation networks* (ICS Report 8901). La Jolla, CA: Institute for Cognitive Science.
- Baba, N. (1989). A new approach for finding the global minimum of error function of neural networks. *Neural Networks*, 2(5), 367-375.
- Baba, N. (1990). A hybrid algorithm for finding the global minimum of error function of neural networks. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, 1 - 585-588.
- Bailey, A.W. (1990). Automatic evolution of neural network architectures. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, 1 - 589-592.
- Battiti, R. (1990). Optimization methods for back-propagation: Automatic parameter tuning and faster convergence. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, 1 - 593-596.
- Becker, S., & le Cun, Y. (1988). Improving the convergence of back-propagation learning with second order methods. In D.S. Touretzky (Ed.), *Neural Information Processing Systems 1*, (pp. 29-37). San Mateo, CA: Morgan Kaufmann Publishers.
- Black, M., & Ilisevich, D.R. (1984). Determinants of military retention: Officer and enlisted personnel. *Fifth Quadrennial Review of Military Compensation*.
- Borack, J.I. (1984). *A framework for integrating alternate military manpower supply methods*. San Diego, CA: Navy Personnel Research and Development Center.
- Brieman, L., Freidman, J., Olshen, C., & Stone, B.M. (1984). *Classification and Regression. Trees*. Wordsworth International Group.
- Carpenter, G.A. (1989). Neural network models for pattern recognition and associative memory. *Neural Networks*, 2(4), 243-258.

- Carter, G.M., Chaiken, J.M., Murray, M.P., & Walker, W.E. (1983). *Conceptual design of an enlisted force management system for the Air Force*. Santa Monica, CA: The RAND Corporation.
- Carter, G.M., Murray, M.P., Arguden, R.Y., Brauner, M.K., Abrahamse, A.F., Greenberg, H., & Skoller, D.L. (1987). *Middle-term loss prediction models for the Air Force's enlisted force management system (R3482-AF)*. Santa Monica, CA: The RAND Corporation.
- Carter, G.M., Skoller, D.L., Perrin, S.E., & Sakai, C.S. (1988). *An enlisted force management system to predict the effects of bonus decision*. Santa Monica, CA: The RAND Corporation.
- Carter, J.P. (1987). Successfully using peak learning rates of 10 (and greater) in back-propagation networks with the heuristic learning algorithm. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987*, 2, II - 645-651.
- Chauvin, Y. (1990). Dynamic behavior of constrained back-propagation networks. In D.S. Touretzky (Ed.), *Neural Information Processing Systems 2* (pp. 642-649). San Mateo, CA: Morgan Kaufmann Publishers.
- Chen, J., & Mars, P. (1990). Stepwise variation methods for accelerating the back-propagation algorithm. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, I - 601-604.
- Cho, S., & Kim, J. (1990). An accelerated learning method with backpropagation. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, D.C., IEEE & INNS*, 1, I - 605-608.
- Cirie, J.A., Miller, J.J., & Sinaiko, H.W. (Eds.) (1981). *Department of Defense and Navy personnel supply models* (ONR TR-9). Arlington, VA: Office of Naval Research.
- Cowin, M., O'Connor, F.E., Sage, L.C., & Johnson, C.R. (1980). *The effects of local economic conditions on Navy enlistment* (NR 170-870). Arlington, VA: Office of Naval Research.
- Curtis, E.W., Borack, J.I., & Wax, S.R. (1987). *Estimating the youth population qualified for military service* (NPRDC-TR-87-32). San Diego, CA: Navy Personnel Research and Development Center.
- Dahl, E.D. (1987). Accelerated learning using the generalized delta rule. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987*, 2, II - 523-530.
- Darken, C., & Moody, J. (1990). Fast adaptive K-means clustering: Some empirical evidence. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 1.
- Defense Advanced Research Projects Agency (1988). *DARPA Neural network study*. Fairfax, VA: AFCFA International Press.
- de Bollivier, M., Galliari, P., & Thiria, S. (1990). Cooperation of neural networks for robust classification. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 113-120.

- DeSieno, D. (1988). Adding a conscience to competitive learning. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988, 1, 1 - 117-124.*
- DeVany, A.S., & Saving, T.R. (1982). Life-cycle job choice and the demand and supply of entry level jobs: Some evidence from the Air Force. *The Review of Economics and Statistics*, LXIV(3), 457-465.
- DeVany, A.S., Saving, T.R., & Shughart, W.F. (1978). *Supply and equilibrium inventory of Air Force enlisted personnel: A simultaneous model of the accession and retention markets incorporating force level constraints.* (AFHRL-TR-78-10, AD-A058 097) Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.
- Dewan, H., & Sontag, E. (1990). Extrapolatory methods for speeding up the backpropagation algorithm. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS, 1, 1 - 613-616.*
- Duda, R.O., & Hart, P.E. (1973). *Pattern classification and scene analysis.* New York: John Wiley & Sons.
- Elman, J.L. (1989). *Representation and structure in connection models* (CRL-8903). University of California at San Diego: Center for Research in Language.
- Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211
- Fahlmann, S. (1988). Faster-learning variations on back-propagation and empirical study. In D. Touretzky (Ed.), *Neural Information Processing Systems 1*, (pp. 38-51). San Mateo, CA: Morgan Kaufmann Publishers.
- Fast, J.C., & Looper, L.T. (1988). *Multiattribute decision modeling techniques: A comparative analysis* (AFHRL-TR-88-3, AD-A201 329). Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.
- Fernandez, R.L., Gotz, G.A., & Bell, R.M. (1985). *The dynamic retention model* (N-2141-MIL). Santa Monica, CA: The RAND Corporation.
- Funahashi, K. (1989). On the approximate realization of continuous mappings by Neural Network. *Neural Networks*, 2(3), 183-192.
- Goldberg, D.E. (1989). *Genetic algorithms.* New York: Addison-Wesley Publishing.
- Gotz, G.A., & McCall, J.J., (1985). *A dynamic retention model for Air Force officers* (R-3028-AF). Santa Monica CA: The RAND Corporation.
- Gotz, G.A., & McCall, J.J. (1980). *Estimating military personnel retention rates: Theory and statistical method* (R-2541-AF). Santa Monica, CA: The RAND Corporation.
- Grossberg, S. (1969). Embedding fields: A theory of learning with physiological implications. *Journal of Mathematical Psychology*, 6, 209-239.

- Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 217-257.
- Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control*. Boston, MA: D. Riedel.
- Grossberg, S. (Ed.). (1986). *Adaptive Brain* (Vols 1-2). New York: Elsevier.
- Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1(1), 17-62.
- Hagiwara, M. (1990). Accelerated backpropagation using unlearning based on a Hebb rule. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington DC, IEEE & INNS*, 1, 1 - 617-620.
- Hanson, S.J., & Pratt, L.Y. (1989). Comparing biases for minimal network construction with back-propagation. In D. S. Touretzky (Ed.), *Neural Information Processing Systems 1* (pp. 177-185). San Mateo, CA: Morgan Kaufmann Publishers.
- Hartigan, J.A. (1975). *Clustering Algorithms*. New York: John Wiley & Sons, Inc.
- Hecht-Nielsen, R. (1987a). Counterpropagation networks. *Applied Optics*, 26(23), 4979-4984.
- Hecht-Nielsen, R. (1987b). Counterpropagation networks. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987*, 2, 11 - 19-32.
- Hecht-Nielsen, R. (1987c). Kolmogorov mapping neural network existence theorem. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987*, 3, 11-13.
- Hefietz, S. (1989). Oral Presentation at *International Joint Conference on Neural Networks, Washington, DC, June 18-22, 1989, IEEE & INNS*, 2, 1 - 1.
- Holland, J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.
- Hosek, J.R., & Petersen, C.E. (1986). *Enlistment choice among different strata of male youth*. Santa Monica, CA: The RAND Corporation.
- Hoskins, J.C. (1989). *Speeding up artificial neural networks in the "real" world* (pre-print). Austin, TX: Microelectronics and Computer Technology Corporation.
- Jacobs, R.A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1, 295-307.

- Kalbfleisch, J.D., & Prentice, R.L. (1980). *The statistical analysis of failure time data*. New York: John Wiley & Sons.
- Kangas, J., Kohonen, T., Leaksnes, J., Simula, O., & Venta, O. (1989). Variants of self-organizing maps. *International Joint Conference on Neural Networks, Washington, DC, June 18-22, 1989, IEEE & INNS*, 2, 11 - 517-522.
- Kimoto, T., Asakana, K., Yoda, M., & Takewaka, M. (1990). Stock market prediction with modular neural networks. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, 1 - 1-6.
- Kohler, D.F. (1988). *Using survivor functions to estimate occupation-specific bonus effect* (R-3348-FMP). Santa Monica, CA: The RAND Corporation.
- Kohonen, T. (1982a). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Kohonen, T. (1982b). Analysis of a simple self-organizing process. *Biological Cybernetics*, 44, 135-140.
- Kohonen, T. (1984). *Self-organization and associative memory* (3rd ed.). New York: Springer-Verlag.
- Kohonen, T. (1988). An introduction to neural computing. *Neural Networks*, 1(1), 3-16.
- Kohonen, T. (1989). Self-Organizing Semantic Maps. *Biological Cybernetics*, 61, 241-254.
- Kohonen, T. (1990). Improved versions of learning vector quantization. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, 1 - 545-550.
- Kollias, S., & Anastassiou, D. (1988). Adaptive training of multilayer neural networks using a least squares technique. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988*, 1, 11-13.
- Kosko, B. (1990). Stochastic competitive learning. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, 1 - 1.
- Kramer, A.H., & Sangiovanni-Vincentelli, A. (1989). Efficient parallel learning algorithms for neural networks. In D. S. Touretzky (Ed.), *Neural Information Processing Systems 1* (pp. 40-48). San Mateo, CA: Morgan Kaufmann Publishers.
- Kruschke, J.K. (1988). Creating local and distributed bottlenecks in hidden layers of back-propagation networks. In D. Touretzky, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School, June 17-26, 1988, Carnegie Mellon University* (pp. 120-126). San Mateo, CA: Morgan Kaufmann Publishers.

- Lakhani, H. (1987). *Bonuses, training costs, and quit rates: A three stage least squares approach* (TR 734). Alexandria, VA: Manpower and Personnel Research Laboratory, U.S. Army Research Institute for the Behavioral and Social Sciences.
- Lakhani, H., Gilroy, C., & Capps, C. (1984). *Logistic microdata model of first term Army reenlistment*. (MPPRG working paper 84-2). Alexandria, VA: Manpower and Personnel Research Laboratory, U.S. Army Research Institute for the Behavioral and Social Sciences.
- Lapedes, A., & Farber, R. (1987). *Nonlinear signal processing using neural networks: Prediction and system modelling* (Preprint LA-UR-87-2662). Los Alamos, NM: Los Alamos National Laboratory.
- Leamer, E.E. (1978). *Specification Searches*. New York: John Wiley & Sons.
- Lincoln, W.P., & Strzypek, J. (1990). Synergy of clustering multiple back propagation networks. In D.S. Touretzky (Ed.), *Neural Information Processing Systems 2* (pp. 650-657). San Mateo, CA: Morgan Kaufmann Publishers.
- Lippman, R.P. (1987). An introduction to computing with neural networks. *IEEE ASSP Magazine*, 4, 4-22.
- Looper, L.T., (1979). *Career area rotation model (CAROM): Historical overview of technique and utilization* (AFHRL-TR-78-97, AD-A068 820). Brooks AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.
- Looper, L.T., & Beswick, C.A. (1980). *Recruiting resource and goal allocation decision model*. (AFHRL-TR-79-55, AD-A080 747). Brooks AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.
- Matyas, J. (1965). Random optimization. *Automation and Remote Control*, 26, 246-253.
- McGarrity, J.P. (1988). *An Air Force selection model for the Armed Forces Health Professions Scholarship Program* (AFHRL-SR-88-24, AD-A200 563). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- McInerney, J.M., Haines, K.G., Biafore, S., & Hecht-Nielsen, R. (1989). Back propagation error surfaces can have local minima. *International Joint Conference on Neural Networks, Washington, DC, June 18-22, 1989, IEEE and INNS*, 2, 11-627.
- Michelson, W.T., & Rydell, P.C. (1989). *Aggregate dynamic model (ADAM) for Air Force enlisted personnel: Technical documentation* (N-3020/2-AF). Santa Monica, CA: The RAND Corporation.
- Minai, A., & Williams, R. (1990). Accelerating of back-propagation through learning rate and momentum adaptation. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, 1 - 677-679.

- Moody, J., & Darken, C. (1988). Learning with localized receptive fields. In D. Touretski, G. Hinton, & T. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Models Summer School, June 17-26, 1988, Carnegie Mellon University* (p. 133). San Mateo, CA: Morgan Kaufmann Publishers.
- Morgan, N., & Bourlard, H. (1990). Generalization and parameter estimation in feedforward nets: Some experiments. In D. S. Touretzky (Ed.), *Neural Information Processing Systems 2* (pp. 630-637). San Mateo, CA: Morgan Kaufmann Publishers.
- Movellan, J. (1990). Error functions to improve noise resistance and generalization in backpropagation networks. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington DC, IEEE & INNS*, 1, 1 - 557-560.
- Mozer, M.C., & Smolenski, P. (1989). Skeletonization: A technique for trimming the fat from a network via relevance assessment. In D. S. Touretzky (Ed.), *Neural Information Processing Systems 1* (pp. 107-115). San Mateo, CA: Morgan Kaufmann Publishers.
- Ohnishi, A., Okamoto, A., & Sugie, N. (1990). Selective presentation of learning samples for efficient learning in multi-layer perceptron. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, 1 - 688-691.
- Orvis, B.R., & Gahart, M.T. (1985). *Relationships of enlistment intention and market survey information to enlistment in active duty military* (N-2292-MIL). Santa Monica, CA: The RAND Corporation.
- Orvis, B.R., & Gahart, M.T. (1989). *Quality-based analysis capability for national youth surveys: Development, application and implications for policy* (R-3675-FMP). Santa Monica, CA: The RAND Corporation.
- Orvis, B.R., Gahart, M.T., & Hosek, J.R. (1989). *Predicting enlistment for recruiting market segments* (N-2852-FMP). Santa Monica, CA: The RAND Corporation.
- Palmieri, F., & Shaw, S. (1990). Fast training of multilayer perceptrons using multilinear parametrization. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington DC*, 1, 1 - 696-699.
- Patrikar, A., & Provence, J. (1990). Learning by local variations. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC*, 1, 1 - 700-703.
- Rezgui, A., & Tepedelenlioglu, N. (1990). The effect of the slope of the activation function on the back propagation algorithm. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington DC, IEEE & INNS*, 1, 1 - 707-710.
- Rueter, F.H., Kosy, D.W., Caicco, G.E., Laidlaw, C.D., & Looper, L.T., (1981). *Integrated Simulation Evaluation Model Prototype (ISEM-P) of the Air Force manpower personnel system: Overview and sensitivity analysis* (AFHRL-TR-81-15, AD-A103 065). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Rumelhart, D.E. (1990). *Brain style computation: Neural networks and connectionist AI*. Paper presented at TIMS/ORSA Joint National Meeting, Las Vegas, May 7-9, 1990.

- Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1984). *Learning internal representations by error propagation* (ICS report 8506). San Diego, CA: University of California, San Diego, Institute for Cognitive Science.
- Rumelhart, D.E., & McClelland, J.L. (Eds.). (1986). *Parallel Distributed Processing, Explorations in the Microstructure of Cognitions* (Vols 1-2). Cambridge, MA: The MIT Press.
- Samad, T. (1990). Backpropagation improvements based on heuristic arguments. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, IEEE & INNS*, 1, 1 - 565-568.
- Saving, T.R., & Stone, B.S. (1982). *Analysis of the impact of "people programs" upon retention of enlisted personnel in the Air Force* (Contract No. F41689-81-C-0063). Randolph AFB, TX: AFMPC/MPCYA, Final Report.
- Saving, T.R., Stone, B.S., Looper, L.T., & Taylor, J. (1985). *Retention of Air Force enlisted personnel: An empirical examination*. (AFHRL-TP-85-6, AD-A158 091). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Scalero, R., & Tepedelenioglu, N. (1990). A fast training algorithm for neural networks. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC*, 1, 1 - 715-717.
- Siegel, B.S., & Borack, J.I. (1981). *An econometric model of Navy enlistment behavior* (NPRDC TN 81-16). San Diego, CA: Navy Personnel Research and Development Center.
- Sietsma, J., & Dow, R.J.F. (1988). Neural net pruning - why and how. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988*, 1, 1 - 325-333.
- Simpson, P.K. (1990). *Artificial neural systems: Foundations, paradigms, applications, and implementations*. Elmsford, NY: Pergamon Press.
- Singhal, J., & Wu, L. (1989). Training multilayer perceptrons with the extended Kalman algorithm. In D. Touretzky (Ed.), *Neural Information Processing Systems 1* (pp. 133-130). San Mateo, CA: Morgan Kaufmann Publishers.
- Smith, A.D., Sylvester, S., & Villa, C. (1989). *Army reenlistment models*. (MDA903-87-C-0590). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Solis, F.J., & Wets, J.B. (1981). Minimization by random search techniques. *Mathematics of Operations Research*, 6, 19-30.
- Specht, D.F. (1988). Probabilistic neural networks for classification, mapping, or associative memory. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988*, 1, 1 - 525-532.
- Specht, D.F. (1990). Probabilistic neural networks. *Neural Networks*, 3(1), 109-118.

- Spoehr, K.T., & Lehmkuhle, S.W. (1982). *Signal detection theory, visual information processing*. New York, NY: Freeman.
- Spirkovska, L., & Reid, M. (1990). Connectivity strategies for HONN's applied to pattern recognition. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, 1 - 21-26.
- Stone, B.S., Looper, L.T., & McGarrity, J.P. (1990b). *Validation and reestimation of an Air Force reenlistment analysis model* (AFHRL-TP-89-55, AD-A218 794). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Stone, B.S., & Saving, T.R. (1983). *Air Force prior service market analysis* (Contract No. F49642-82-CO54). Washington, DC: AF/MPXOA.
- Stone, B.S., Saving, T.R., Turner, K.L., Looper, L.T., & Engquist, S.K. (1991). *A simultaneous estimation model of Air Force accession and retention* (AL-TR-1991-0063, AD-A240 064). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Research Division.
- Stone, B.S., Wortman, R.A., & Looper, L.T. (1989). *Air Force Retention Analysis Package: Users manual* (AFHRL-TP-89-10, AD-A212 767). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Stornetta, W.S., & Huberman, B.A. (1990). An improved three-layer, back propagation algorithm. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987*, 2, II - 637-643.
- Syllogistics Inc., & RRC, Inc. (1989). *Epps VI.0 design manual*. Springfield VA: Syllogistics Inc., for the Analysis Division and Planning Division, Directorate of Personnel Plans, Washington, DC.
- Terza, J.V., & Warren, R.S. (1986). *U.S. Army reenlistment and extension by occupation: A reduced-form trinomial probit approach* (DAAG29-81-D-0100, Delivery Order 1748). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Verdugo, N., & Berliant, K. (1988). *Estimating the Army's prime recruiting market* (ARI-TR-832). Alexandria, VA: Manpower and Personnel Research Laboratory, U.S. Army Research Institute for the Behavioral and Social Sciences.
- Vogl, T.P., Mangis, J.K., Rigler, A.K., Zink, W.T., & Alkon, D.L. (1988). Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, 59, 257-263.
- Warner, J.T., & Goldberg, M.S. (1983). The influence of non-pecuniary factors on labor supply: The case of Navy enlisted personnel. *The Review of Economics and Statistics*, XX(Y), 26-35.
- Wasserman, P.D. (1989). *Neural computing, theory and practice*. New York: Van Nostrand Reinhold.

- Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences* (Ph.D. thesis). Cambridge MA: Harvard University Committee on Applied Mathematics.
- Whitley, D., & Starkweather, T. (1990). Optimizing small neural networks using a distributed genetic algorithm. In M. Caudill (Ed.), *International Joint Conference on Neural Networks, January 15-19, 1990, Washington, DC, 1*, 1 - 207-209.
- Widrow, B. (1990). Paper presented at *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*.
- Wiggins, V. (1990a). *Neural network applications literature review*. Informal technical information prepared for Air Force Human Resources Laboratory, Brooks AFB, TX (Contract No. F41689-88-D-0251, Task 29). Bryan, TX: RRC, Inc.
- Wiggins, V., Looper, L., & Engquist, S. (1991). *Neural network: A primer* (AL-TP-1991-0011, AD-A235 920). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Manpower and Personnel Research Division.